# ONE WAY FANOVA USING PENALIZED SPLINES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Matthías Kormáksson

August 2007

## ABSTRACT

There are several methods available for smoothing scatter-plots. One interesting method involves using mixed model techniques that can be shown to be equivalent to the penalized splines method. In order to analyze certain functional data sets, we propose an extension of this mixed model approach that involves the smoothing of several scatter-plots simultaneously. More precisely, we show how one can estimate the mean profiles of functional data that have one grouping factor by fitting a single mixed model. The underlying mixed model will then be used to set up a hypotheses testing scheme for doing one way functional analysis of variance, FANOVA. In doing so, we will establish an interesting connection between the one way FANOVA problem and the problem of testing whether variance components from certain mixed models are zero. Finally we will propose a method for doing multiple comparisons in the functional setting, again using the underlying mixed model from the fitting criteria. The proposed methods are then demonstrated through an analysis of a typical functional data set.

## BIOGRAPHICAL SKETCH

Matthías Kormáksson was born May 6th 1981 in Sauðárkrókur, a small town in the northern most part of Iceland. At the age of four he moved to Reykjavik and has since lived in the surroundings of the capital. In the years 1997-2001 he was a student at the secondary school Menntaskólinn í Reykjavík, where he majored in Math and Physics. In the Fall of 2001, he enrolled in the Math department at the University of Iceland, where he graduated in the spring of 2004 with a B.Sc. degree. After doing three years of pure Math, Matthías realized that he wanted to do something more applied and decided to pursue a Ph.D. in Statistics. He was accepted into the Ph.D. program at the Department of Statistical Science, Cornell University in the Fall of 2004, and is expected to finish in the Spring of 2009.

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my adviser, professor James Booth, for his countless suggestions and insightful hints while conducting my research and writing this thesis. His intuition on the material has been inspirational and has greatly helped me bridge the gap between my exceedingly analytical mind and my less utilized intuitive mind. I believe that bridging this gap is a crucial part of the process of becoming a good Statistician.

Secondly, I would like to thank the second member of my special committee, professor Robert Strawderman, for the time and effort he put into comments for improving my thesis. His sincere commitment not only to sit on my committee, but also lend a helping hand whenever needed, is greatly appreciated.

I am also very grateful to assistant professor Giles Hooker, for his devotion to proofread my thesis. His comments and suggestions were extremely useful. In addition, he helped me gain more insight into the world of Functional data analysis through the fda-course he taught in the Spring of 2007.

Finally, I would like to thank my fellow student and friend, Michael Grabchak, for sharing different perspectives on research-related problems. I would also like to express my appreciation of the mutual support we provide whenever we feel like we have reached a dead end in our research. Surprisingly, this happens often, and knowing one is not alone with these thoughts is very comforting.

While doing this research, I was partially supported by funding from my adviser's National Science Foundation grant (NSF-DMS-04-05543). I am sincerely grateful for this generous support.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# Chapter 1

# Introduction

Functional data analysis is a field within statistics that has received increasing attention in the last few decades. The number of functional data sets arising in various scientific fields is growing fast, driving the need for statistical techniques for inference. The pioneering work of Grace Wahba drew attention to the importance of the field starting in the early 70's. Since then, several statisticians have contributed to the field, making it to what it is today. A broad overview of available methods was given in Ramsay and Silverman (1997). Their book provides excellent perspectives on the field and has recently been updated, Ramsay and Silverman (2005), to keep up with the fast development in the last decade.

Despite all of the great work, there is still a large demand for new statistical tools for analyzing the vast number of existing functional data sets. One interesting problem, functional analysis of variance (FANOVA), is to compare samples of curves across different groups. More precisely, we wish to test the hypothesis of all mean curves, across groups, being equal. Ramsay and Silverman (1997) discuss this problem, but seem to emphasize the point wise FANOVA; i.e. testing $f_1(t) = \ldots = f_g(t)$ at all time points $t$. The nature of that problem is different than that of testing $f_1 = \ldots = f_g$, over the whole range of $x$-values. Some interesting work on the latter problem, includes Cuevas et al. (2004), but they derive asymptotic results based on a test statistic that is a functional of the sample trajectories. They give a thorough review of the literature on FANOVA in their paper. In this thesis, we will propose a method for testing the null hypothesis

$$H_0 : f_1 = \ldots = f_g, \tag{1.1}$$

using methods purely from mixed model literature. This approach is different from earlier work on FANOVA, which, for the most part, takes place within some functional space, such as $L_2$.

In order to approach the FANOVA problem in this manner, we will suggest a mixed model of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \varepsilon, \tag{1.2}$$

that consists of a fixed polynomial regression part, a random penalized spline part, and random noise. The response vector, $\mathbf{Y}$, contains all the discretized subject profiles within all the groups evaluated at the same design points $(t_k)_{k=1}^r$. The design matrices are set up accordingly to account for all of the profiles. We will extend the penalized splines as BLUPs method, see Brumback et al. (1999), to multiple curves, providing us with a way of smoothing all of the mean profiles simultaneously using a simple mixed model. This idea of extending individual curve smoothing to multiple curves in a mixed model setting has been pointed out by Brumback and Rice (1998), but they proposed a mixed model that extends smoothing spline procedures from one curve to a sample of curves. The advantage of using penalized splines over smoothing splines is the significant computational savings. This results from the fact that penalized splines are low-rank smoothers, as defined in Hastie (1996), and involve only inversions of $K \times K$ matrices, where $K$ is the number of knots. As a brief historical note, Kimeldorf and Wahba (1970) and Wahba (1978) derived smoothing splines as a Bayes estimate when smoothing individual curves. Their setting was in continuous time, but Silverman (1985) later came up with the discrete analogy, and then Speed (1991) pointed out the connection between the smoothing splines and BLUPs of a certain mixed model. This idea of a mixed model representation of smoothing splines was then adapted

to the penalized spline smoother of Eilers and Marx (1996) by Brumback et al. (1999).

In this thesis, we will show how the model in (1.2) can be fitted in a very standard way using REML criteria to obtain estimates of our mean functions. The fitting criteria involves matrix calculations in high dimensions, but we will show how the structure of the design matrices can be used to our advantage to reduce the dimensionality of our problem substantially. The aforementioned method we propose for testing the hypothesis (1.1) is a hypotheses testing scheme consisting of three steps. Two of these steps will involve testing whether certain variance components arising from the mixed model (1.2) are equal to zero. For testing these hypotheses we will use the elegant results of Crainiceanu and Ruppert (2004). The second step of our testing scheme, is particularly note worthy. It essentially tests whether the mean curves across groups have the same kind of smoothing; or, in other words, whether the mean curve differences can be modeled using a fixed polynomial regression model. If the curves are indeed equal, we will fail to reject the *step 2.* hypothesis. We can then set up a contrast and test to see if that contrast is zero using linear regression. Upon rejecting the null hypothesis in (1.1) we might want to ask where the differences between the curves lie, and this raises the question of multiple comparisons. What is a natural way of comparing several curves over a continuum of x-values? We will propose a method involving a construction of simultaneous confidence bands about mean curve differences. This author believes that this is a natural way of testing for quantitative differences between curves. But this procedure will fail if the differences between the curves are not of quantitative nature. It is, for example, conceivable that we might reject the null hypothesis, (1.1), but still detect no differences between curves when

constructing simultaneous confidence bands. This phenomena will be addressed with an example in the discussions in chapter 6. To illustrate the ideas presented in this thesis we will run a data analysis on a typical functional data set.

We will also look at whether averaging the raw profiles within groups will result in loss of information in terms of estimating and making inferences about the mean curves. We will compare this mean model to the full model and show that if the number of curves per group, $n_i$, is large the mean model is preferred. Fitting the mean model makes the estimating procedure substantially simpler, and allows for the use of standard software. We will also state and prove some consistency results but show why one cannot consistently estimate the variance components of the mixed model (1.2) using the REML estimates, if the number of groups and/or knots is bounded.

# Chapter 2

# Penalized splines as BLUPS

## 2.1 Estimation and prediction

In this section we will discuss the estimation of $\boldsymbol{\beta}$ and prediction of $\mathbf{u}$ in a mixed model of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{2.1}$$

where

$$\text{Cov}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

The random vectors $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ are both considered multivariate normal with mean $\mathbf{0}$ and covariance matrices $\mathbf{G} = \sigma_u^2\mathbf{I}$ and $\mathbf{R} = \sigma_\varepsilon^2\mathbf{I}$ respectively.

The BLUE and BLUP of the two parameter vectors will be presented but they depend on the parameters of the covariance matrices $\mathbf{G}$ and $\mathbf{R}$ and hence the estimation of those parameters will also be treated. By plugging in the estimates of $\mathbf{G}$ and $\mathbf{R}$, the estimated BLUPs (EBLUPs) are obtained. For more detailed discussion on these estimators see Ruppert et al. (2003).

When $\mathbf{Y}$ has a multivariate normal distribution, the BLUE of $\boldsymbol{\beta}$ and the BLUP of $\mathbf{u}$ can be shown to equal

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \\ \tilde{\mathbf{u}} &= \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \end{aligned} \tag{2.2}$$

where $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$. These formulas are originally due to Henderson (1950).

As mentioned above the BLUPs in (2.2) depend on the parameters of the covariance matrices $\mathbf{G}$ and $\mathbf{R}$. In general these matrices can have various structures, but

in the mixed model (2.1) the parameters in question are simply the two variance components $\sigma_\varepsilon^2$ and $\sigma_u^2$. There are various methods available for estimating these parameters and the two most popular ones are the *maximum likelihood* (ML) and the *restricted maximum likelihood* (REML) algorithms. The difference between the two methods is that the REML algorithm takes into account the degrees of freedom for the fixed effects in (2.1). Details on the derivation of the REML criteria, originally due to Patterson and Thompson (1971), can be found in Searle et al. (1992). In this thesis the REML algorithm will be used throughout. The *restricted log-likelihood* for the model is

$$\begin{aligned} l_R &= -\frac{1}{2}[\log|\mathbf{V}| + \mathbf{y}^T\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right)\mathbf{y} \\ &\quad - \log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|] - \frac{n}{2}\log(2\pi) \end{aligned} \tag{2.3}$$

and the REML algorithm involves maximizing $l_R$ over the parameters of $\mathbf{V}$, i.e. maximizing $l_R$ over the variance component pair $(\sigma_\varepsilon^2, \sigma_u^2)$. As we will see in Chapter 3 it turns out that with some simplifications, this maximization problem can be reduced to a one dimensional problem.

After obtaining the REML estimates $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_u^2$ and setting $\hat{\mathbf{R}} = \hat{\sigma}_\varepsilon^2\mathbf{I}$ and $\hat{\mathbf{G}} = \hat{\sigma}_u^2\mathbf{I}$, we can use the BLUP formulas (2.2) to obtain the EBLUP estimates:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \{\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\}^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} &= \hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned} \tag{2.4}$$

where $\hat{\mathbf{V}} = \mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}^T + \hat{\mathbf{R}}$.

## 2.2 Smoothing with a mixed model approach

Consider the problem of smoothing a scatterplot in the plane. The scatterplot consists of data points $(x_i, y_i)$, $i = 1, \ldots, r$, and the goal is to estimate an unspecified

smooth function $f$ from the model:

$$y_i = f(x_i) + \varepsilon_i, \quad E(\varepsilon_i) = 0, \tag{2.5}$$

where $\varepsilon_i$ is regarded as random noise in the data.

There are various methods available for estimating the function $f$. Some popular methods include *penalized splines*, *local polynomial regression* and some series-based methods that use, for example, *Fourier* or *wavelet* bases. The methodology used throughout this thesis is sought from the mixed model literature. The estimate of $f$ can be shown to equal the best linear unbiased predictor, BLUP of a certain mixed model. Ruppert et al. (2003) give an excellent description of the mixed model criteria, and show that the estimate is equivalent to the one obtained using *penalized splines*.

The idea is to model the function $f$ using a spline model. In what follows the quadratic spline will be used:

$$f(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \sum_{m=1}^{K} u_m (x_i - \tau_m)_+^2 \tag{2.6}$$

where $\tau_m$ are called knots and are spread out over the range of the $x$-values. Note that this is exactly the spline term of the *penalized spline* model. A typical penalization term for (2.6) would be

$$\lambda^2 \sum_{m=1}^{K} u_m^2$$

In the mixed model approach there is however no penalization term but instead the coefficients of (2.6), $\beta_0, \beta_1, \beta_2$ are treated as fixed and $u_1, \ldots, u_K$ treated as random. Then using standard mixed model techniques the target function $f$ can be estimated.

More formally, let

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix}$$

and define

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_r & x_r^2 \end{bmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} (x_1-\tau_1)_+^2 & \dots & (x_1-\tau_K)_+^2 \\ \vdots & \ddots & \vdots \\ (x_r-\tau_1)_+^2 & \dots & (x_r-\tau_K)_+^2 \end{bmatrix} \tag{2.7}$$

then with the assumption of normality of $\mathbf{u}$ and $\boldsymbol{\varepsilon}$, the model (2.5) can be formulated as a mixed model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{2.8}$$

with

$$\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{R}) \quad \text{and} \quad \mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$$

where $\mathbf{G} = \sigma_u^2 \mathbf{I}_r$ and $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_r$ respectively.

It follows that the parameters of the mixed model (2.8) with design matrices (2.7) can be estimated to obtain the EBLUPs $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ of the form (2.4). Plugging these into (2.8) results in the fitted values

$$\hat{\mathbf{f}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$$

and this leads us to the estimate of our target function $f$ at an arbitrary x-value:

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \sum_{m=1}^{K} \hat{u}_m (x - \tau_m)_+^2$$

## 2.3 Generalization of the mixed model representation

In this section a more general mixed model will be presented. The basic idea behind this model is to smooth several scatter plots simultaneously. The model can be

applied to a certain class of functional data and is more tractable than some of the methods currently used in functional data analysis, since it falls directly into the mixed model framework. The mixed model (2.1) is a special case of this general model.

Consider a functional response $\mathbf{y}_{ij}(t)$ measured on different subjects $j$ within different groups $i$. An example of such a functional response could be heat measurements taken over a one year time period at different weather stations in different weather regions, see Canadian Weather data in Ramsay and Silverman (1997). The functional data is of course discrete but the underlying function is assumed to be continuous and in some sense well behaved. Now the typical methodology in functional data analysis involves fitting a curve through the data points for each subject separately. Then each curve is treated as a known functional response and one can proceed to perform functional analysis of variance (FANOVA). In the mixed model representation below a mean curve for each group will be estimated using only the raw discrete data from all the subjects. This makes the whole estimation procedure a lot simpler in the sense that there is no need for pre-smoothing all together.

The data considered in this thesis have the same discretization. This means that the response, $\mathbf{y}_{ij}(t)$, is observed at the same points, $t = t_1, \ldots, t_r$, for all subjects, $j = 1, \ldots, n_i$, in all groups, $i = 1, \ldots, g$. The total number of subjects will be denoted $N = \sum_{i=1}^{g} n_i$. For each subject $j$ in group $i$ there corresponds a scatter plot $(t_k, \mathbf{y}_{ijk})$, $k = 1, \ldots, r$, and the assumption is that for each group there is an unspecified function $f_i$ such that

$$\mathbf{y}_{ij}(t_k) = f_i(t_k) + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

Often in practice, especially when one is dealing with time series data, this as-

sumption of independence between $\varepsilon_{ijk}$ and $\varepsilon_{ijk'}$ is not going to be valid. The main reason for still using the model above is the simple nature of it. In fact, the substantial simplifications in the fitting procedure, discussed in chapter 3, rely on the diagonal structure of the error covariance matrix. A possible remedy, is to estimate the covariance matrix and then do a Cholesky decomposition, in order to maintain the diagonal structure. This idea will be discussed in more detail in chapter 6. It is worth noting though, that in terms of fitting the data, the assumption of independence is not going to affect the results to any extent. It is only when one starts doing inference, (chapter 4), that the assumption might impact the results.

As in section 2.2 the functions $f_i$ will be modeled using a quadratic spline function:

$$f_i(t_k) = \beta_{0i} + \beta_{1i}t_k + \beta_{2i}t_k^2 + \sum_{m=1}^{K} u_{mi}(t_k - \tau_m)_+^2 \tag{2.9}$$

where the coefficients, $\beta_{0i}, \beta_{1i}, \beta_{2i}$, are considered fixed and the coefficients, $u_{1i}, \ldots, u_{Ki}$, iid normally distributed with mean 0 and variance $\sigma_u^2$.

This can again all be formulated as a mixed model just as in the one curve smoothing problem described in section 2.2. The only difference in this setting lies in the design matrices $\mathbf{X}$ and $\mathbf{Z}$. Define

$$\boldsymbol{\beta}_i = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_i = \begin{bmatrix} u_{1i} \\ \vdots \\ u_{Ki} \end{bmatrix}$$

and let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_g^T)^T$ and $\mathbf{u} = (\mathbf{u}_1^T, \ldots, \mathbf{u}_g^T)^T$. Define further

$$\mathbf{B} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_r & t_r^2 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} (t_1-\tau_1)_+^2 & \dots & (t_1-\tau_K)_+^2 \\ \vdots & \ddots & \vdots \\ (t_r-\tau_1)_+^2 & \dots & (t_r-\tau_K)_+^2 \end{bmatrix}$$

and let

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_1} \otimes \mathbf{B} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \otimes \mathbf{B} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_g} \otimes \mathbf{B} \end{bmatrix} \tag{2.10}$$

and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{1}_{n_1} \otimes \mathbf{C} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \otimes \mathbf{C} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_g} \otimes \mathbf{C} \end{bmatrix} \tag{2.11}$$

Note that $\mathbf{B}$ and $\mathbf{C}$ are equivalent to the design matrices (2.7) in the one curve smoothing mixed model. Here they serve as building blocks in our new design matrices and each pair occurring in $\mathbf{X}$ and $\mathbf{Z}$ respectively corresponds to a scatterplot for a certain subject.

Let $\mathbf{y}_{ij} = (y_{ij}(t_1), \dots, y_{ij}(t_r))^T$, $\mathbf{y}_i = (\mathbf{y}_{i1}^T, \dots, \mathbf{y}_{in_i}^T)^T$ and $\mathbf{Y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_g^T)^T$. Let $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \dots, \varepsilon_{ijr})^T$, $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}^T, \dots, \boldsymbol{\varepsilon}_{in_i}^T)^T$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^T, \dots, \boldsymbol{\varepsilon}_g^T)^T$. With these definitions, and the design matrices in (2.10) and (2.11), the mixed model representation of our raw functional data is obtained:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{2.12}$$

with

$$\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{R}) \quad \text{and} \quad \mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$$

where $\mathbf{G} = \sigma_u^2 \mathbf{I}_{gK}$ and $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_{Nr}$.

Now the restricted log-likelihood (2.3) with the more general design matrices above can be maximized through the REML algorithm to obtain estimates for $\sigma_u^2$ and $\sigma_\varepsilon^2$. Then we can obtain the EBLUPs (2.4) to obtain estimates of the functions $f_1, \ldots, f_g$ in (2.9):

$$\hat{f}_i(t) = \hat{\beta}_{0i} + \hat{\beta}_{1i} t + \hat{\beta}_{2i} t^2 + \sum_{m=1}^{K} \hat{u}_{mi} (t - \tau_m)_+^2$$

at an arbitrary t.

There is however a problem that arises when dealing directly with the restricted log-likelihood (2.3). The problem lies in the sizes of the new design matrices $\mathbf{X}$ and $\mathbf{Z}$. The number of rows in both of these matrices is $Nr$ as opposed to just $r$ in (2.7). Hence dealing with the log-likelihood (2.3) directly can lead to computational problems if the total number of subjects $N$ is large. Fortunately there is a way around this problem. It turns out that the block-diagonal structure of $\mathbf{X}$ and $\mathbf{Z}$, where the blocks have a further repetitive structure, can be used to simplify calculations in the REML algorithm substantially. This simplification will be discussed in detail in Chapter 3.

In the functional data analysis literature the following model is more common than the one presented in this section

$$\mathbf{y}(t) = \boldsymbol{\mu}(t) + \boldsymbol{\gamma}_i(t) + \varepsilon(t) \tag{2.13}$$

where $\boldsymbol{\mu}(t)$ represents the overall mean curve, $\boldsymbol{\gamma}_i(t)$ represents the deviation of the mean curve for group $i$ from the overall mean curve and $\varepsilon(t)$ is the random error curve. It is easy to reformulate the model discussed above to an equivalent model of the specific form (2.13). Let $\bar{\beta}_{l\cdot} = \frac{1}{g} \sum_{i=1}^{g} \beta_{li}$, for $l = 0, 1, 2$ and $\bar{u}_{m\cdot} = \frac{1}{g} \sum_{i=1}^{g} u_{mi}$ for $m = 1, \ldots, K$. Let also $\beta_{li}^* = (\beta_{li} - \bar{\beta}_{l\cdot})$, for $l = 0, 1, 2$, and $u_{mi}^* = (u_{mi} - \bar{u}_{m\cdot})$,

for $m = 1, \ldots, K$. Then with

$$\boldsymbol{\mu}(t) = \bar{\beta}_{0\cdot} + \bar{\beta}_{1\cdot} t + \bar{\beta}_{2\cdot} t^2 + \sum_{m=1}^{K} \bar{u}_{m\cdot} (t - \tau_m)_+^2$$

and

$$\boldsymbol{\gamma}_i(t) = \beta_{0i}^* + \beta_{1i}^* t + \beta_{2i}^* t^2 + \sum_{m=1}^{K} u_{mi}^* (t - \tau_m)_+^2$$

we can formulate our model from above in the desired way

$$\mathbf{y}_{ij}(t_k) = \boldsymbol{\mu}(t_k) + \boldsymbol{\gamma}_i(t_k) + \varepsilon_{ijk}. \tag{2.14}$$

# Chapter 3

# Estimation

In this chapter we will discuss the general problem of estimating the variance components in a simple mixed model of the form (2.1). The maximization of the restricted log-likelihood (2.3) will be treated, and we will discuss how the two dimensional maximization problem can be reduced to a one dimensional one. This will result in a simpler form of the EBLUPs given in section 2.1.

As was mentioned in section 2.3 the main problem, that arises when dealing directly with the restricted log-likelihood for the functional mixed model (2.12) is the high dimensionality of the design matrices. Working directly with these high dimensional matrices can lead to serious computational issues. We will therefore introduce simplified formulas for the restricted log-likelihood and it's first and second derivatives in that setting. The structure of the design matrices will be used to show that the dimensions of the matrices involved are in fact the same as the ones in the REML-estimation for the one curve smoothing of section 2.2.

## 3.1 The general REML setup

The restricted log-likelihood for a general mixed model of the form (2.1) is:

$$\begin{aligned} l_R(\sigma_\varepsilon^2, \sigma_u^2) &= -\frac{1}{2}[\log|\mathbf{V}| + \mathbf{y}^T\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right)\mathbf{y} \\ &\quad + \log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|] - \frac{n}{2}\log(2\pi) \end{aligned}$$

where

$$\mathbf{V} = \mathrm{Cov}\,(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} = \sigma_u^2\mathbf{Z}\mathbf{Z}^T + \sigma_\varepsilon^2\mathbf{I}$$

Ruppert et al. (2003) show that by letting $\alpha = \sigma_\varepsilon^2/\sigma_u^2$ and defining the matrices

$$\mathbf{A}(\alpha) = \alpha\mathbf{I} \quad \text{and} \quad \mathbf{\Psi}(\alpha) = \sigma_\varepsilon^2\mathbf{V}^{-1}$$

the restricted log-likelihood takes on the form:

$$\begin{aligned} l_R(\sigma_\varepsilon^2, \alpha) &= -\frac{1}{2}[(n-p)\log(\sigma_\varepsilon^2) + \{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\alpha)\}^T\mathbf{\Psi}(\alpha)\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\alpha)\}/\sigma_\varepsilon^2 \\ &\quad + \log|\mathbf{I} + \mathbf{Z}^T\mathbf{Z}\mathbf{A}(\alpha)^{-1}| + \log|\mathbf{X}^T\mathbf{\Psi}(\alpha)\mathbf{X}|] - \frac{n}{2}\log(2\pi), \end{aligned} \tag{3.1}$$

where $p = 3g$ is the number of columns in the design matrix $\mathbf{X}$, and the following set of computing formulas hold:

$$\begin{aligned} \boldsymbol{\beta}(\alpha) &= \{\mathbf{X}^T\mathbf{\Psi}(\alpha)\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{\Psi}(\alpha)\mathbf{y}, \\ \mathbf{\Psi}(\alpha) &= \mathbf{I} - \mathbf{Z}\{\mathbf{A}(\alpha) + \mathbf{Z}^T\mathbf{Z}\}^{-1}\mathbf{Z}^T. \end{aligned} \tag{3.2}$$

They furthermore state that for a fixed $\alpha$, $l_R(\sigma_\varepsilon^2, \alpha)$ is maximized over $\sigma_\varepsilon^2 > 0$ by

$$\hat{\sigma}_\varepsilon^2(\alpha) = \{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\alpha)\}^T\mathbf{\Psi}(\alpha)\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\alpha)\}/(n-p).$$

By plugging this estimate into $l_R(\sigma_\varepsilon^2, \alpha)$ the following expression for the restricted log-likelihood is obtained (up to a constant):

$$\begin{aligned} l_R(\alpha) &= -\frac{n-p}{2}\log\Big(\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\alpha)\}^T\mathbf{\Psi}(\alpha)\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\alpha)\}/(n-p)\Big) \\ &\quad -\frac{1}{2}\log|\mathbf{I} + \mathbf{Z}^T\mathbf{Z}\mathbf{A}(\alpha)^{-1}| \\ &\quad -\frac{1}{2}\log|\mathbf{X}^T\mathbf{\Psi}(\alpha)\mathbf{X}|, \end{aligned} \tag{3.3}$$

and all that remains is maximization of (3.3) over $\alpha \in \mathbb{R}_+$.

After obtaining estimates for the variance components $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_u^2$ and the ratio $\hat{\alpha} = \hat{\sigma}_\varepsilon^2/\hat{\sigma}_u^2$ we can use the EBLUP formulas (2.4) for $\boldsymbol{\beta}$ and $\mathbf{u}$, and the computing formulas (3.2) above to obtain the simplified EBLUPs:

(recall $\boldsymbol{\Psi}(\alpha) = \sigma_\varepsilon^2 \mathbf{V}^{-1}$)

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \{\mathbf{X}^T \boldsymbol{\Psi}(\hat{\alpha}) \mathbf{X}\}^{-1} \mathbf{X}^T \boldsymbol{\Psi}(\hat{\alpha}) \mathbf{y}, \\ \hat{\mathbf{u}} &= \mathbf{Z}^T \boldsymbol{\Psi}(\hat{\alpha})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/\hat{\alpha}. \end{aligned} \tag{3.4}$$

## 3.2 REML for the functional mixed model

In this section we will discuss some simplifications of the restricted log-likelihood (3.3) when we are dealing with the functional mixed model discussed in section 2.3. As mentioned before, working directly with the expressions in (3.3) can be computationally infeasible when the design matrices take on the forms in (2.10) and (2.11). The matrix $\boldsymbol{\Psi}(\alpha)$, for example, is an $Nr \times Nr$ matrix and can be huge when the total number of curves $N$ is large.

It is possible to simplify things by noting that $\mathbf{X}$ and $\mathbf{Z}$ are block-diagonal with blocks

$$\mathbf{X}_i = \mathbf{1}_{n_i} \otimes \mathbf{B}, \quad \text{and} \quad \mathbf{Z}_i = \mathbf{1}_{n_i} \otimes \mathbf{C}$$

respectively. By defining

$$\mathbf{A}_i(\alpha) = \alpha \mathbf{I}_K,$$

for all $i = 1, \ldots, g$, where $K$ is the number of knots, it is easily verified that

$$\boldsymbol{\Psi}(\alpha) = \mathbf{I}_{Nr} - \mathbf{Z}\{\mathbf{A}(\alpha) + \mathbf{Z}^T\mathbf{Z}\}^{-1}\mathbf{Z}^T$$

is also block-diagonal with blocks

$$\boldsymbol{\Psi}_i(\alpha) = \mathbf{I}_{n_i r} - \mathbf{Z}_i\{\mathbf{A}_i(\alpha) + \mathbf{Z}_i^T\mathbf{Z}_i\}^{-1}\mathbf{Z}_i^T.$$

Finally the fixed parameter vector

$$\boldsymbol{\beta}(\alpha) = \{\mathbf{X}^T \boldsymbol{\Psi}(\alpha) \mathbf{X}\}^{-1} \mathbf{X}^T \boldsymbol{\Psi}(\alpha) \mathbf{y}$$

can be divided into sub-vectors $\boldsymbol{\beta}(\alpha) = (\boldsymbol{\beta}_1^T(\alpha), \ldots, \boldsymbol{\beta}_g^T(\alpha))^T$ where

$$\boldsymbol{\beta}_i(\alpha) = \{\mathbf{X}_i^T \boldsymbol{\Psi}_i(\alpha) \mathbf{X}_i\}^{-1} \mathbf{X}_i^T \boldsymbol{\Psi}_i(\alpha) \mathbf{y}_i \tag{3.5}$$

corresponds to the fixed parameters for group $i$.

Now it is evident that the restricted log-likelihood (3.3) becomes:

$$\begin{aligned} l_R(\alpha) &= -\frac{n-p}{2} \log \sum_{i=1}^{g} \Big( \{\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i(\alpha)\}^T \boldsymbol{\Psi}_i(\alpha) \{\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i(\alpha)\}/(n-p) \Big) \\ &\quad -\frac{1}{2} \sum_{i=1}^{g} \log |\mathbf{I} + \mathbf{Z}_i^T \mathbf{Z}_i \mathbf{A}_i(\alpha)^{-1}| \\ &\quad -\frac{1}{2} \sum_{i=1}^{g} \log |\mathbf{X}_i^T \boldsymbol{\Psi}_i(\alpha) \mathbf{X}_i|, \end{aligned} \tag{3.6}$$

and upon finding the MLEs, $\hat{\sigma}_\varepsilon^2$, $\hat{\sigma}_u^2$ and $\hat{\alpha}$, the above facts can be used to simplify (3.4) to obtain EBLUPs for the parameters in each group $i$:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i &= \{\mathbf{X}_i^T \boldsymbol{\Psi}_i(\hat{\alpha}) \mathbf{X}_i\}^{-1} \mathbf{X}_i^T \boldsymbol{\Psi}_i(\hat{\alpha}) \mathbf{y}_i \\ \hat{\mathbf{u}}_i &= \mathbf{Z}_i^T \boldsymbol{\Psi}_i(\hat{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i)/\hat{\alpha} \end{aligned} \tag{3.7}$$

Finally plugging these estimates into the spline models (2.9) we obtain the estimates of the groups mean curves for our functional data:

$$\hat{f}_i(t) = \hat{\beta}_{0i} + \hat{\beta}_{1i} t + \hat{\beta}_{2i} t^2 + \sum_{m=1}^{K} \hat{u}_{mi} (t - \tau_m)_+^2. \tag{3.8}$$

But even with the simplifications described above, it is still not feasible to deal with (3.6) directly. The dimension of $\boldsymbol{\Psi}_i(\alpha)$ is $n_i r$ which can still be computationally prohibitive if the number of subjects $n_i$ in a specific group $i$ is large. It turns out that with the repetitive structure of $\mathbf{X}_i = \mathbf{1}_{n_i} \otimes \mathbf{B}$ and $\mathbf{Z}_i = \mathbf{1}_{n_i} \otimes \mathbf{C}$ further simplification of the log-likelihood (3.6) and its first and second derivatives is possible. In fact the matrices $\mathbf{B}$ and $\mathbf{C}$ determine the dimensions of the matrices involved in our simplified formulas just as in the one curve fitting scheme discussed in section 2.2.

## 3.3 Computing formulas for estimation

In this section computing formulas for the restricted log-likelihood (3.6) and its first and second derivatives are stated. These formulas can be used directly to implement the Newton-Raphson Algorithm for obtaining the REML estimates of the variance components in the functional mixed model (2.12). Then we will state simplified computing formulas for the EBLUPs $\hat{\boldsymbol{\beta}}_i$ and $\hat{\mathbf{u}}_i$ from (3.7). Details on the derivation of these formulas can be found in Appendix A.

Define

$$\begin{aligned}
\mathbf{A}_{1i} &:= \frac{1}{n_i}\mathbf{C}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T \\
d\mathbf{A}_{1i} &:= \frac{d}{d\alpha}\mathbf{A}_{1i} \\
&= -\frac{1}{n_i^2}\mathbf{C}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T \\
d^2\mathbf{A}_{1i} &:= \frac{d^2}{d\alpha^2}\mathbf{A}_{1i} \\
&= \frac{2}{n_i^3}\mathbf{C}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T. \qquad (3.9)
\end{aligned}$$

Then define

$$\begin{aligned}
\mathbf{W}_i &:= \mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i \\
&= n_i\mathbf{B}^T(\mathbf{I}-n_i\mathbf{A}_{1i})\mathbf{B} \\
d\mathbf{W}_i &:= \frac{d}{d\alpha}\mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i \\
&= -n_i^2\mathbf{B}^T d\mathbf{A}_{1i}\mathbf{B} \\
d^2\mathbf{W}_i &:= \frac{d^2}{d\alpha^2}\mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i \\
&= -n_i^2\mathbf{B}^T d^2\mathbf{A}_{1i}\mathbf{B}. \qquad (3.10)
\end{aligned}$$

Further define

$$
\begin{aligned}
\mathbf{A}_{2i} &:= \mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i}) \\
d\mathbf{A}_{2i} &:= \frac{d}{d\alpha}\mathbf{A}_{2i} \\
&= -n_i\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T d\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i}) \\
d^2\mathbf{A}_{2i} &:= \frac{d^2}{d\alpha^2}\mathbf{A}_{2i} \\
&= -n_i\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T\Big\{d^2\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i}) - 2n_i d\mathbf{A}_{1i}d\mathbf{A}_{2i}\Big\}, \qquad (3.11)
\end{aligned}
$$

and finally

$$
\begin{aligned}
S_i(\mathbf{y},\alpha) &:= \sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}) \\
&\qquad +n_i\{\bar{\mathbf{y}}_{\mathbf{i}.} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{\mathbf{i}.} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\} \\
&= \sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}) \\
&\qquad +n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}),
\end{aligned}
$$

which has derivatives

$$
\begin{aligned}
S_i'(\mathbf{y},\alpha) &= -2n_i(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
&\quad -n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}),
\end{aligned}
$$

and

$$
\begin{aligned}
S_i''(\mathbf{y},\alpha) &= -2n_i(n_i d^2\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
&\quad +4n_i(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
&\quad +2n_i(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
&\quad -n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d^2\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}).
\end{aligned}
$$

With the definitions above at hand, the log-likelihood becomes:

$$
\begin{aligned}
l_R(\alpha) &= -\frac{n-p}{2}\log\sum_{i=1}^{g}\frac{1}{n-p}S_i(\mathbf{y},\alpha) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{I}_K+\frac{n_i}{\alpha}\mathbf{C}^T\mathbf{C}| \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\log|n_i\mathbf{B}^T(\mathbf{I}_r-n_i\mathbf{A}_{1i})\mathbf{B}|. \qquad (3.12)
\end{aligned}
$$

The differentiated log-likelihood becomes:

$$
\begin{aligned}
dl_R(\alpha) &= -\frac{n-p}{2}\left(\frac{\sum_{i=1}^{g}S_i'(\mathbf{y},\alpha)}{\sum_{i=1}^{g}S_i(\mathbf{y},\alpha)}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\left(\frac{1}{n_i}\mathrm{tr}\Big\{(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I}_K)^{-1}\Big\}-\frac{K}{\alpha}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\left(\mathrm{tr}\Big\{\mathbf{W}_i^{-1}(d\mathbf{W}_i)\Big\}\right), \qquad (3.13)
\end{aligned}
$$

and the twice differentiated log-likelihood becomes:

$$
\begin{aligned}
d^2l_R(\alpha) &= -\frac{n-p}{2}\left(\frac{\left(\sum_{i=1}^{g}S_i''(\mathbf{y},\alpha)\right)\left(\sum_{i=1}^{g}S_i(\mathbf{y},\alpha)\right)-\left(\sum_{i=1}^{g}S_i'(\mathbf{y},\alpha)\right)^2}{\left(\sum_{i=1}^{g}S_i(\mathbf{y},\alpha)\right)^2}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\left(-\frac{1}{n_i^2}\mathrm{tr}\Big\{(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I}_K)^{-1}(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I}_K)^{-1}\Big\}+\frac{K}{\alpha^2}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\mathrm{tr}\Big\{-\mathbf{W}_i^{-1}(d\mathbf{W}_i)\mathbf{W}_i^{-1}(d\mathbf{W}_i)+\mathbf{W}_i^{-1}(d^2\mathbf{W}_i)\Big\}, \qquad (3.14)
\end{aligned}
$$

The EBLUPs from (3.7) have the following simplified forms:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_i &= \left(\mathbf{B}^T\left(\mathbf{I}_r-n_i\mathbf{A}_{1i}\right)\mathbf{B}\right)^{-1}\mathbf{B}^T\left(\mathbf{I}_r-n_i\mathbf{A}_{1i}\right)\bar{\mathbf{y}}_{i\cdot}, \\
\hat{\mathbf{u}}_i &= \frac{n_i}{\hat{\alpha}}\mathbf{C}^T\left(\mathbf{I}_r-n_i\mathbf{A}_{1i}\right)(\bar{\mathbf{y}}_{i\cdot}-\mathbf{B}\hat{\boldsymbol{\beta}}_i). \qquad (3.15)
\end{aligned}
$$

If we take a closer look at the formulas (3.12), (3.13) and (3.14) used to implement the Newton-Raphson algorithm we see that the matrices $\mathbf{B}$ and $\mathbf{C}$, that are equivalent to the design matrices from the one curve smoothing mixed model, are the building units. The same applies to the EBLUP estimates in (3.15). Thus

what we have really accomplished here is reducing the $Nr$ dimensional problem to an $r$ dimensional one just like in the one curve smoothing scheme. Another thing to notice when we look at the first part of the restricted log-likelihood in (3.12) is that the statistic $\{\mathbf{S}^2, (\bar{\mathbf{y}}_{i.})_{i=1}^{g}\}$ is a sufficient statistic, where

$$\mathbf{S}^2 = \frac{1}{n-p}\sum_{i=1}^{g}\sum_{j=1}^{n_0}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}).$$

This can be readily seen by the factorization theorem.

## 3.4 The mean profile mixed model

So far we have been focusing on the functional mixed model

$$\mathbf{y}_{ij}(t_k) = \beta_{0i} + \beta_{1i}t_k + \beta_{2i}t_k^2 + \sum_{m=1}^{K} u_{mi}(t_k - \tau_m)_+^2 + \varepsilon_{ijk}, \tag{3.16}$$

where $\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. This model is based on the full data in the sense that we use the discretized data for each subject, $j = 1, \ldots, n_i$, within each group, $i = 1, \ldots, g$. Now consider averaging the discrete data over the subjects in each group. Then we get a model of the form

$$\bar{\mathbf{y}}_{i.}(t_k) = \beta_{0i} + \beta_{1i}t_k + \beta_{2i}t_k^2 + \sum_{m=1}^{K} u_{mi}(t_k - \tau_m)_+^2 + \bar{\varepsilon}_{i.k}, \tag{3.17}$$

where the errors are all independent, but not necessarily identical, with $\bar{\varepsilon}_{i.k} \sim N(0, \sigma_\varepsilon^2/n_i)$ for all $i = 1, \ldots, g$. The random slopes, $u_{mi}$, are as before considered iid normal, with mean 0 and variance $\sigma_u^2$.

Note that for each $i$, independently, (3.17) is a mixed model just like the full model (3.16). The error variance of the new mean profile model is $\sigma_\varepsilon^2/n_i$, which differs from the error variance of the full model by a scaling factor. On the other hand, the variance component, $\sigma_u^2$, the regression parameters, $\beta_{0i}, \beta_{1i}, \beta_{2i}$, and the

random effects, $u_{1i}, \ldots, u_{Ki}$, are identical across both models. A natural question to ask is whether we lose any information by simply using the mean profiles. In this section we will derive the restricted log-likelihood for the mean model and compare it to that of the full model. When estimating parameters, it is standard practice to plug in the REML estimates of the variance components without taking into account the extra variability due to estimation. We will show that following that procedure one will obtain similar results using both models when the number of curves per group, $n_i$, is big. Unfortunately one cannot establish any consistency results for the REML estimates but we will nonetheless state some consistency results for the error variance and prove them in appendix $B$. We will also explain why we cannot obtain the aforementioned consistency results for the REML estimates.

The mean profile model (3.17) can be formulated in matrix notation as

$$\bar{\mathbf{Y}} = \mathbf{X}_g \boldsymbol{\beta} + \mathbf{Z}_g \mathbf{u} + \bar{\boldsymbol{\varepsilon}}, \tag{3.18}$$

where $\bar{\mathbf{Y}} = \left(\bar{\mathbf{Y}}_{\mathbf{1}\cdot}^T, \ldots, \bar{\mathbf{Y}}_{\mathbf{g}\cdot}^T\right)^T$ is just the group average profiles stacked together, with

$$\bar{\mathbf{Y}}_{\mathbf{i}\cdot} = \left(\bar{\mathbf{y}}_{\mathbf{i.1}}, \ldots, \bar{\mathbf{y}}_{\mathbf{i.r}}\right)^T,$$

for $i = 1, \ldots, g$. Similarly the error vector, $\bar{\boldsymbol{\varepsilon}}$, is just the averaged error vectors within the groups stacked together. The design matrices are

$$\mathbf{X}_g = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{B} \end{bmatrix}$$

and

$$\mathbf{Z}_g = \begin{bmatrix} \mathbf{C} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{C} \end{bmatrix},$$

where the $\mathbf{B}$'s and $\mathbf{C}$'s occur $g$ times on the diagonals.

Note that the vectors $\bar{\mathbf{Y}}_{\mathbf{1}\cdot}, \dots, \bar{\mathbf{Y}}_{\mathbf{g}\cdot}$ are independent and each follow a mixed model of the form

$$\bar{\mathbf{Y}}_{\mathbf{i}\cdot} = \mathbf{B}\boldsymbol{\beta}_i + \mathbf{C}\mathbf{u}_i + \bar{\boldsymbol{\varepsilon}}_{i\cdot}, \tag{3.19}$$

where

$$\mathbf{u}_i \sim \mathbf{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_K) \quad \text{and} \quad \bar{\boldsymbol{\varepsilon}}_{i\cdot} \sim \mathbf{N}(\mathbf{0}, (\sigma_\varepsilon^2/n_i)\mathbf{I}_r).$$

Hence the likelihood of the mean vector, $\bar{\mathbf{Y}}$, is just the product of the likelihoods of each of the group mean vectors $\bar{\mathbf{Y}}_{\mathbf{i}\cdot}$.

We follow the argument in section 3.1 to derive the restricted log-likelihood of (3.19) for each $i$. The restricted log-likelihood takes on the form

$$\begin{aligned} l_R^i(\sigma_i^2, \sigma_u^2) &= -\frac{1}{2}[\log|\mathbf{V}_i| + \bar{\mathbf{y}}_{\mathbf{i}\cdot}^T \left(\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{B}\left(\mathbf{B}^T\mathbf{V}_i^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{V}_i^{-1}\right)\bar{\mathbf{y}}_{\mathbf{i}\cdot} \\ &\quad + \log|\mathbf{B}^T\mathbf{V}_i^{-1}\mathbf{B}|] - \frac{r}{2}\log(2\pi), \end{aligned}$$

where

$$\mathbf{V}_i = \mathrm{Cov}\,(\bar{\mathbf{y}}_{\mathbf{i}\cdot}) = \sigma_u^2\mathbf{C}\mathbf{C}^T + \sigma_i^2\mathbf{I}_r,$$

with $\sigma_i^2 := \sigma_\varepsilon^2/n_i$. We define similarly to what we did in section 3.1

$$\mathbf{A}_i^*(\alpha) = (\alpha/n_i)\mathbf{I} \quad \text{and} \quad \boldsymbol{\Psi}_i^*(\alpha) = (\sigma_\varepsilon^2/n_i)\mathbf{V}_i^{-1}$$

and note that $\alpha/n_i = (\sigma_\varepsilon^2/n_i)/\sigma_u^2$ is simply the variance components ratio for the mixed model (3.19). For each group $i$, we obtain the following form for the

restricted log-likelihood

$$\begin{aligned} l_R^i(\sigma_i^2,\alpha) &= -\frac{1}{2}\Big[(r-3)\log(\sigma_i^2) + \{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}^T\boldsymbol{\Psi}_i^*(\alpha)\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}/\sigma_i^2 \\ &\quad + \log|\mathbf{I} + \mathbf{C}^T\mathbf{C}\mathbf{A}_i^*(\alpha)^{-1}| + \log|\mathbf{B}^T\boldsymbol{\Psi}_i^*(\alpha)\mathbf{B}|\Big] - \frac{r}{2}\log(2\pi) \end{aligned} \tag{3.20}$$

and the following holds

$$\begin{aligned} \boldsymbol{\Psi}_i^*(\alpha) &= \mathbf{I}_r - \mathbf{C}\{\mathbf{A}_i^*(\alpha) + \mathbf{C}^T\mathbf{C}\}^{-1}\mathbf{C}^T \\ &= \mathbf{I}_r - n_i\mathbf{A}_{1i}, \\ \boldsymbol{\beta}_i^*(\alpha) &= \{\mathbf{B}^T\boldsymbol{\Psi}_i^*(\alpha)\mathbf{B}\}^{-1}\mathbf{B}^T\boldsymbol{\Psi}_i^*(\alpha)\bar{\mathbf{y}}_{\mathbf{i}\cdot}, \\ &= \left(\mathbf{B}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\mathbf{B}\right)^{-1}\mathbf{B}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\bar{\mathbf{y}}_{\mathbf{i}\cdot}. \end{aligned}$$

Notice that $\boldsymbol{\beta}_i^*(\alpha) = \boldsymbol{\beta}_i(\alpha)$, from (3.15). For a fixed $\alpha$, (3.20) is maximized at

$$\hat{\sigma}_\varepsilon^2 = n_i\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}^T\boldsymbol{\Psi}_i^*(\alpha)\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}/(r-3), \tag{3.21}$$

which is a $\sqrt{r}$-consistent estimator of $\sigma_\varepsilon^2$, at the true value of $\alpha$, for each $i = 1,\ldots,g$, as $r \to \infty$ (see proof in appendix $B$).

Since the mean profiles are independent, we obtain from (3.20) the joint restricted log-likelihood of $\bar{\mathbf{Y}}_{\mathbf{1}\cdot},\ldots,\bar{\mathbf{Y}}_{\mathbf{g}\cdot}$

$$\begin{aligned} l_R(\sigma_\varepsilon^2,\alpha) &= -\frac{1}{2}\Big[(r-3)\sum_{i=1}^{g}\log(\sigma_i^2) + gr\log(2\pi) \\ &\quad + \sum_{i=1}^{g}\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}^T\boldsymbol{\Psi}_i^*(\alpha)\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}/\sigma_i^2 \\ &\quad + \sum_{i=1}^{g}\log|\mathbf{I} + \mathbf{C}^T\mathbf{C}\mathbf{A}^*(\alpha)^{-1}| + \sum_{i=1}^{g}\log|\mathbf{B}^T\boldsymbol{\Psi}_i^*(\alpha)\mathbf{B}|\Big] \end{aligned}$$

$$
\begin{aligned}
= \; & -\frac{1}{2}\Big[(gr-p)\log(\sigma_\varepsilon^2) + gr\log(2\pi) - (gr-p)\log n_i \\
& + \sum_{i=1}^{g} n_i\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}/\sigma_\varepsilon^2 \\
& + \sum_{i=1}^{g}\log|\mathbf{I} + \frac{n_i}{\alpha}\mathbf{C}^T\mathbf{C}| + \sum_{i=1}^{g}\log|\mathbf{B}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\mathbf{B}|\Big].
\end{aligned}
$$

Again we use the argument that for a fixed $\alpha$, $l_R(\sigma_\varepsilon^2, \alpha)$ is maximized over $\sigma_\varepsilon^2 > 0$ with

$$
\hat{\sigma}_\varepsilon^2 = \sum_{i=1}^{g} n_i\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}/(gr-p)
$$

and by plugging into $l_R(\sigma_\varepsilon^2, \alpha)$ we obtain up to a constant with respect to $\alpha$:

$$
\begin{aligned}
l_R(\alpha) \; = \; & -\frac{gr-p}{2}\log\sum_{i=1}^{g}\frac{n_i}{gr-p}\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\} \\
& -\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{I}_K + \frac{n_i}{\alpha}\mathbf{C}^T\mathbf{C}| \\
& -\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{B}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\mathbf{B}| \qquad (3.22)
\end{aligned}
$$

and now what everything boils down to is maximizing (3.22) with respect to $\alpha$. Upon finding a maximizer $\hat{\alpha}$ we obtain the EBLUPs for the parameters of the mean model

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_i &= \left(\mathbf{B}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\mathbf{B}\right)^{-1}\mathbf{B}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})\,\bar{\mathbf{y}}_{i\cdot}, \\
\hat{\mathbf{u}}_i &= \frac{n_i}{\hat{\alpha}}\mathbf{C}^T(\mathbf{I}_r - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i\cdot} - \mathbf{B}\hat{\boldsymbol{\beta}}_i). \qquad (3.23)
\end{aligned}
$$

We notice immediately that these formulas are identical to the EBLUP formulas for the full model, (3.15). Of course the two models will result in different estimates of $\alpha$ and $\sigma_\varepsilon^2$ and hence different parameter estimates. But as we will now argue, when the number of curves per group, $n_i$, is large we will expect the resulting parameter

estimates to be very similar using either criteria. With some manipulations, the EBLUP formulas in (3.23) can be written in the following manner

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{B}^T(\mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C} + \frac{\hat{\alpha}}{n_i}\mathbf{I})^{-1}\mathbf{C}^T)\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C} + \frac{\hat{\alpha}}{n_i}\mathbf{I})^{-1}\mathbf{C}^T)\bar{\mathbf{y}}_{i\cdot}$$

$$\hat{\mathbf{u}}_i = (\mathbf{C}^T\mathbf{C} + \frac{\hat{\alpha}}{n_i}\mathbf{I})^{-1}\mathbf{C}^T(\bar{\mathbf{y}}_{i\cdot} - \mathbf{B}\hat{\boldsymbol{\beta}}_i)$$

and we see that if $n_i$ is large, the parameter estimates should be robust to different values of $\alpha$. Hence using either the mean model or the full model should result in similar parameter estimates and thus similar mean curve estimates. When constructing confidence bands around the mean curves one needs the following fact, which holds at the true value of $\alpha$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_i(\alpha) - \boldsymbol{\beta}_i \\ \hat{\mathbf{u}}_i(\alpha) - \mathbf{u}_i \end{bmatrix} \sim \mathbf{N}\left\{\mathbf{0}, \frac{\sigma_\varepsilon^2}{n_i}\left(\mathbf{S}^T\mathbf{S} + \frac{\alpha}{n_i}\mathbf{D}\right)^{-1}\right\}, \tag{3.24}$$

where $\mathbf{S} = [\mathbf{B} \quad \mathbf{C}]$ and $\mathbf{D} = \text{diag}(0, 0, 0, 1, \ldots, 1)$. The three zeros and the $K$ ones of $\mathbf{D}$ correspond to the fixed and random effects of the quadratic spline. The construction of confidence bands involves simulating from this distribution with the true $\sigma_\varepsilon^2$ and $\alpha$ replaced by their REML estimates. This will result in approximate confidence bands around the mean curves. We will discuss this in detail in chapter 4 but the main thing to notice here is that again if $n_i$ is large we expect the results not to depend heavily on the estimate of $\alpha$. In fact we have the following asymptotic result

$$\sqrt{n_i}\begin{bmatrix} \hat{\boldsymbol{\beta}}_i(\alpha) - \boldsymbol{\beta}_i \\ \hat{\mathbf{u}}_i(\alpha) - \mathbf{u}_i \end{bmatrix} \longrightarrow \mathbf{N}\left\{\mathbf{0}, \sigma_\varepsilon^2(\mathbf{S}^T\mathbf{S})^{-1}\right\} \tag{3.25}$$

as $n_i \to \infty$ and we see that the asymptotic covariance does not depend on $\alpha$. When simulating approximate confidence bands based on (3.24), it is standard

practice to simply plug in the REML estimates of the variances and ignore the extra variability due to the estimation. Following that procedure, we recommend fitting the mean model when the number of curves is large since the results are expected to be similar even for different values of $\alpha$. If $n_i$ is large one could also use the result in (3.25) and ignore the estimated value of $\alpha$ all together. The reason we choose the mean model instead of the full model is the substantial savings in computation time. Also fitting the mean model will allow one to use packages such as proc mixed of SAS and lme of S-plus. At the true value of $\alpha$, it turns out that both in the mean and the full model, $\hat{\sigma}^2_\varepsilon(\alpha)$ is consistent as $r, n_i$ go to infinity. These facts are proven in appendix $B$. But the problem is that one cannot consistently estimate $\alpha$ and hence one cannot use these consistency results directly. As we will discuss below, these consistency results for the error variance could become relevant as the number of groups, $g$, or the number of knots, $K$, go to infinity. On the other hand, as $n_i \to \infty$, one can consistently estimate the error variance using the sample variance

$$\hat{\sigma}^2_\varepsilon = \frac{1}{n-p} \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}) \tag{3.26}$$

and hence an alternative is to fit the mean model and use the REML estimate of $\alpha$ but the moment estimate of $\sigma^2_\varepsilon$. The consistency of (3.26) will be shown in appendix $B$.

As mentioned above we cannot consistently estimate $\alpha$. This is due to the fact that the number of groups, $g$, and the number of knots, $K$, is fixed and finite and hence the number of random parameters, $u_{11}, \ldots, u_{Kg}$, is $gK < \infty$. That means that there has to be an upper bound on the information about the variance component $\sigma^2_u$ and hence $\alpha$. Moreover, in practice we have even less information since the random parameters are not directly observed. It would be

interesting however to see if we obtain consistency by letting the number of groups, $g$, or number of knots, $K$, go to infinity as well. Of course in practice letting the number of groups go to infinity would typically not be realistic. However, letting the number of knots increase with the number of sample points, $r$, is realistic.

# Chapter 4

# Hypotheses testing

In One Way ANOVA one is interested in testing whether or not there are differences between the means of several groups. If difference is declared then multiple comparison techniques can be used to find out where the differences lie. The Functional ANOVA (FANOVA) setting is a little bit more complicated since we are dealing with mean curves rather than scalar means. In the multiple comparison problem for example it is not clear what exactly it means for two curves to be different. In fact, the two curves could coincide on parts of the curves but differ substantially on other parts. In the first section of this chapter we will propose a hypotheses testing scheme for the problem of testing whether the mean curves $f_1, \ldots, f_g$, introduced in section 2.3, are identical across groups; i.e. testing the null-hypothesis

$$H_0 : f_1 = \ldots = f_g$$

This hypothesis is of course equivalent to the null-hypothesis

$$H_0 : \boldsymbol{\gamma}_i = 0, \text{ all i}$$

arising from the equivalent model in (2.14)

$$\mathbf{y}_{ij}(t_k) = \boldsymbol{\mu}(t_k) + \boldsymbol{\gamma}_i(t_k) + \varepsilon_{ijk} \tag{4.1}$$

The hypotheses testing scheme proposed consists of 3 steps, and is to serve as a systematic procedure that deals with the one way FANOVA problem. The first two steps involve testing whether certain variance components arising from the mixed model (2.12) are equal to 0. The third step is a well known and simple hypothesis from multiple regression. Both the balanced case when the number of

subjects in each group are equal and the unbalanced case when they are not will be considered. The second section deals with the problem of multiple comparison in the functional setting.

## 4.1 Likelihood ratio test for the variance component in a simple mixed model

Before describing the 3 steps of the hypotheses testing scheme, consider the problem of testing $H_0 : \sigma_u^2 = 0$ in the mixed model,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon},$$

with

$$\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}) \quad \text{and} \quad \mathbf{u} \sim \mathbf{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}).$$

Crainiceanu and Ruppert (2004) deal with this problem in detail. The main result is given in the following theorem

**Theorem 4.1.1** *Let $\mu_{s,n}$ be the $K$ eigenvalues of the $K \times K$ matrices $\mathbf{Z}^T\mathbf{Z} - \mathbf{Z}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}$, then the restricted likelihood ratio test statistic ($RLRT_n$) for testing the hypothesis $H_0 : \sigma_u^2 = 0$ vs. the alternative $H_a : \sigma_u^2 > 0$ has the following property*

$$RLRT_n \stackrel{D}{=} \sup_{\lambda \geq 0} \left[ (n-p) \log \left\{ 1 + \frac{N_n(\lambda)}{D_n(\lambda)} \right\} - \sum_{s=1}^{K} \log(1 + \lambda \mu_{s,n}) \right]$$

*where $N_n(\lambda)$ and $D_n(\lambda)$ are defined as follows*

$$\begin{aligned} N_n(\lambda) &= \sum_{s=1}^{K} \frac{\lambda \mu_{s,n}}{1 + \lambda \mu_{s,n}} w_s^2, \\ D_n(\lambda) &= \sum_{s=1}^{K} \frac{w_s^2}{1 + \lambda \mu_{s,n}} + \sum_{s=K+1}^{n-p} w_s^2. \end{aligned}$$

*and $w_s$ for $s = 1, \ldots, n-p$ are independent $N(0,1)$.*

In their paper they provide a simple algorithm for simulating the null finite sample distribution of the likelihood ratio test statistic (LRT). The algorithm for simulating the null finite sample distribution of the RLRT can be obtained by direct analogy and follows these 6 steps:

*Step 1.* Define a grid $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_m$ of possible values for $\lambda$.

*Step 2.* Simulate $K$ independent $\chi^2_1$ random variables $w_1^2, \ldots, w_K^2$.

*Step 3.* Independently of step 2, simulate $X_{n,K,p} = \sum_{s=K+1}^{n-p} w_s^2$ with a $\chi^2_{n-p-K}$ distribution.

*Step 4.* For every grid point $\lambda_i$ compute

$$\begin{aligned} N_n(\lambda_i) &= \sum_{s=1}^{K} \frac{\lambda_i \mu_{s,n}}{1 + \lambda_i \mu_{s,n}} w_s^2, \\ D_n(\lambda_i) &= \sum_{s=1}^{K} \frac{w_s^2}{1 + \lambda_i \mu_{s,n}} + X_{n,K,p}. \end{aligned}$$

*Step 5.* Compute

$$RLRT_n = \max_{\lambda_i} \left[ (n-p) \log \left\{ 1 + \frac{N_n(\lambda_i)}{D_n(\lambda_i)} \right\} - \sum_{s=1}^{K} \log(1 + \lambda_i \mu_{s,n}) \right]$$

*Step 6.* Repeat steps $1 - 5$ until the desired number of simulations is achieved.

## 4.2 Hypotheses testing scheme

In this section we start off by stating the three hypotheses steps for testing

$$H_0 : f_1 = \ldots = f_g$$

Then, in the two subsections, we will explain the details of each step for both the balanced and the unbalanced case. The underlying model is the mixed model (2.12) from section 2.3.

***Step 1.*** The first hypothesis concerns testing whether the quadratic spline basis part of

$$f_i(t) = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \sum_{r=1}^{K} u_{ri}(t - \tau_r)_+^2$$

is equal to 0 for all $i$; that is

$$H_0 : \mathbf{u}_1 = \ldots = \mathbf{u}_g = \mathbf{0}. \tag{4.2}$$

This hypothesis implies that a fixed multiple regression model is sufficient to model the mean curves. If the hypothesis is rejected one should proceed to *step 2* below. Otherwise one should jump to *step 3*.

***Step 2.*** If the hypothesis in *step 1* is rejected, it is clear that a fixed model is not sufficient for modeling the mean curves. Then another hypothesis of interest is to test whether or not the mean curves across groups $i$ have the same kind of smoothing. In other words we want to test whether the quadratic spline basis parts of

$$f_i(t_k) = \beta_{0i} + \beta_{1i}t_k + \beta_{2i}t_k^2 + \sum_{m=1}^{K} u_{mi}(t_k - \tau_m)_+^2$$

are equal for all $i$,

$$H_0 : \mathbf{u}_1 = \ldots = \mathbf{u}_g \tag{4.3}$$

Notice that this hypothesis involves testing whether all mean curve differences can be modeled using fixed multiple regression models. If this hypothesis is rejected then clearly there are differences between the mean curves. One can then skip *step 3* and start considering some multiple comparisons of interest. If on the other hand this hypothesis is not rejected one should proceed to *step 3*.

***Step 3.*** If either one of the hypotheses in *step 1* or *step 2* is not rejected, it is indicating that the random effects $\mathbf{u}_i$ are equal for all $i$. Then it is evident that the

final hypothesis should involve testing whether the fixed effects are equal across groups,

$$H_0 : \boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_g \tag{4.4}$$

If difference is not declared it means that the mean curves $f_1, \ldots, f_g$ are all equal. If on the other hand difference is declared one should consider multiple comparison to find out where the differences lie. Multiple comparison techniques in the functional setting are discussed in section (4.3).

It might seem like *step 1* is redundant in the three step hypotheses scheme above and one could do with simply *step 2* and *step 3*. We retain *step 1* for two reasons. Firstly, it is not hard to implement the test, and secondly if we fail to reject the hypothesis it is indicating that the mean curves can be modeled with a simple multiple regression model. Hence smoothing the data using mixed model methods is unnecessary. Standard parametric regression methods are sufficient for comparing the groups.

### 4.2.1 Balanced case

In this section we show in detail how each of the three steps above are followed through in the balanced case. We assume the number of subjects per group is $n_0 := n_1 = \ldots = n_g$.

***Step 1.*** The hypothesis test (4.2) is equivalent to testing whether $\sigma_u^2 = 0$ and can be followed through by using the algorithm from Crainiceanu and Ruppert (2004) described in section 4.1 above. The eigenvalues of the block-diagonal matrix

$$\begin{aligned} & \mathbf{Z}^T\mathbf{Z} - \mathbf{Z}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z} \\ = \; & \text{blockdiag}\{n_i\mathbf{C}^T\mathbf{C} - (n_i\mathbf{C}^T\mathbf{B})(n_i\mathbf{B}^T\mathbf{B})^{-1}(n_i\mathbf{B}^T\mathbf{C})\} \end{aligned}$$

$$= \quad \text{blockdiag}\{n_i \mathbf{C}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{C}\}_{1\leq i\leq g}$$

are easily obtained by computing the eigenvalues of each block separately. They are particularly easy to compute in the balanced case since $n_0 := n_1 = \ldots = n_g$.

***Step 2.*** For testing (4.3) consider the mean model discussed in section 3.4

$$\bar{\mathbf{Y}} = \mathbf{X}_g\boldsymbol{\beta} + \mathbf{Z}_g\mathbf{u} + \bar{\boldsymbol{\varepsilon}} \tag{4.5}$$

where $\bar{\mathbf{Y}} = \left(\bar{\mathbf{Y}}_{1.}^T, \ldots, \bar{\mathbf{Y}}_{g.}^T\right)^T$ and $\bar{\boldsymbol{\varepsilon}} = \left(\bar{\boldsymbol{\varepsilon}}_{1.}^T, \ldots, \bar{\boldsymbol{\varepsilon}}_{g.}^T\right)^T$ and the design matrices are $\mathbf{X}_g = \text{blockdiag}\,(\mathbf{B})_{1\leq i\leq g}$ and $\mathbf{Z}_g = \text{blockdiag}\,(\mathbf{C})_{1\leq i\leq g}$. The idea is now to transform the mixed model (4.5) and test whether the variance component of that transformed mixed model is zero. In the simple case of two groups we could look at the following transformed model

$$\bar{\mathbf{Y}}_{1.} - \bar{\mathbf{Y}}_{2.} = \mathbf{B}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) + \mathbf{C}(\mathbf{u}_1 - \mathbf{u}_2) + (\bar{\boldsymbol{\varepsilon}}_{1.} - \bar{\boldsymbol{\varepsilon}}_{2.})$$

and test whether $\sigma^2_{\mathbf{u}_1-\mathbf{u}_2} = 0$.

In order to follow this through, we need to impose a more general covariance structure on the random effects $\mathbf{u}$ and assume

$$\mathbf{u} \sim \text{N}\Big(\mathbf{0}, \sigma_u^2\{(1-\rho)\mathbf{I}_{gK} + \rho(\mathbf{J}_g \otimes \mathbf{I}_K)\}\Big)$$

where $\rho = \text{corr}(u_{mi_1}, u_{mi_2})$ for all $m = 1, \ldots, K$. Note that if $\rho = 0$ the independence structure is obtained. The random error vector $\bar{\boldsymbol{\varepsilon}}$ is distributed as follows:

$$\bar{\boldsymbol{\varepsilon}} \sim \text{N}(\mathbf{0}, \frac{\sigma_\varepsilon^2}{n_0}\mathbf{I}_{gr})$$

where we recall $n_0 := n_1 = \ldots = n_g$.

Next consider the following transformation:

$$\mathbf{A}\bar{\mathbf{Y}} = \begin{bmatrix} \mathbf{A}_1\bar{\mathbf{Y}} \\ \vdots \\ \mathbf{A}_{g-1}\bar{\mathbf{Y}} \end{bmatrix}$$

where $\mathbf{A}_k = \mathbf{a}_k' \otimes \mathbf{I}_r$, $k = 1, \ldots, g-1$. The following restrictions are put on the $g \times 1$ contrast vectors $\mathbf{a}_k = (a_{k1}, \ldots, a_{kg})'$:

$$\mathbf{a}_i' \mathbf{a}_j = 0 \quad \text{for all } i \neq j$$

and

$$\sum_{i=1}^{g} a_{ki} = 0 \quad \text{for all } k = 1, \ldots, g-1$$

As a convention these linear contrasts are standardized by setting

$$\sum_{i=1}^{g} a_{ki}^2 = 1 \quad \text{for all } k = 1, \ldots, g-1.$$

Under the transformation $\mathbf{A}$ it is easily verified that the right hand side of (4.5) becomes

$$\begin{aligned} \mathbf{A}\bar{\mathbf{Y}} &= \mathbf{A}\mathbf{X}_g\boldsymbol{\beta} + \mathbf{A}\mathbf{Z}_g\mathbf{u} + \mathbf{A}\bar{\boldsymbol{\varepsilon}} \\ &= \mathbf{X}_{g-1}\mathbf{A}^*\boldsymbol{\beta} + \mathbf{Z}_{g-1}\mathbf{A}'\mathbf{u} + \boldsymbol{\varepsilon}^*, \end{aligned} \tag{4.6}$$

where $\mathbf{X}_{g-1} = \text{blockdiag}(\mathbf{B})_{1 \leq i \leq g-1}$ and $\mathbf{Z}_{g-1} = \text{blockdiag}(\mathbf{C})_{1 \leq i \leq g-1}$

$$\boldsymbol{\varepsilon}^* = \begin{bmatrix} \mathbf{A}_1\bar{\boldsymbol{\varepsilon}} \\ \vdots \\ \mathbf{A}_{g-1}\bar{\boldsymbol{\varepsilon}} \end{bmatrix}$$

with $\mathbf{A}_k = \mathbf{a}_k' \otimes \mathbf{I}_r$ as defined above,

$$\mathbf{A}^*\boldsymbol{\beta} = \begin{bmatrix} \mathbf{A}_1^*(\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_g^T)^T \\ \vdots \\ \mathbf{A}_{g-1}^*(\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_g^T)^T \end{bmatrix} \quad \text{and} \quad \mathbf{A}'\mathbf{u} = \begin{bmatrix} \mathbf{A}_1'(\mathbf{u}_1^T, \ldots, \mathbf{u}_g^T)^T \\ \vdots \\ \mathbf{A}_{g-1}'(\mathbf{u}_1^T, \ldots, \mathbf{u}_g^T)^T \end{bmatrix},$$

$\mathbf{A}_k^* = \mathbf{a}_k' \otimes \mathbf{I}_3$ and $\mathbf{A}_k' = \mathbf{a}_k' \otimes \mathbf{I}_K$. The magic number 3 comes from the fact that we are dealing with a quadratic spline basis and $K$ is the number of knots.

As a motivating example consider the case where $g = 4$. Let

$$\begin{aligned}
\mathbf{a}_1' &= \frac{1}{\sqrt{2}}(1, 0, 0, -1), \\
\mathbf{a}_2' &= \frac{1}{\sqrt{2}}(0, 1, -1, 0), \\
\mathbf{a}_3' &= \frac{1}{2}(1, -1, -1, 1).
\end{aligned}$$

Then

$$\mathbf{A}\bar{\mathbf{Y}} = \begin{bmatrix} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_4)/\sqrt{2} \\ (\bar{\mathbf{Y}}_2 - \bar{\mathbf{Y}}_3)/\sqrt{2} \\ (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 - \bar{\mathbf{Y}}_3 + \bar{\mathbf{Y}}_4)/2 \end{bmatrix}, \quad \boldsymbol{\varepsilon}^* = \begin{bmatrix} (\bar{\boldsymbol{\varepsilon}}_1 - \bar{\boldsymbol{\varepsilon}}_4)/\sqrt{2} \\ (\bar{\boldsymbol{\varepsilon}}_2 - \bar{\boldsymbol{\varepsilon}}_3)/\sqrt{2} \\ (\bar{\boldsymbol{\varepsilon}}_1 - \bar{\boldsymbol{\varepsilon}}_2 - \bar{\boldsymbol{\varepsilon}}_3 + \bar{\boldsymbol{\varepsilon}}_4)/2 \end{bmatrix}$$

and

$$\mathbf{A}^*\boldsymbol{\beta} = \begin{bmatrix} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_4)/\sqrt{2} \\ (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_3)/\sqrt{2} \\ (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 - \boldsymbol{\beta}_3 + \boldsymbol{\beta}_4)/2 \end{bmatrix}, \quad \mathbf{A}'\mathbf{u} = \begin{bmatrix} (\mathbf{u}_1 - \mathbf{u}_4)/\sqrt{2} \\ (\mathbf{u}_2 - \mathbf{u}_3)/\sqrt{2} \\ (\mathbf{u}_1 - \mathbf{u}_2 - \mathbf{u}_3 + \mathbf{u}_4)/2 \end{bmatrix}.$$

It is now clear that the model in (4.6) is simply a mixed model with response $\mathbf{A}\bar{\mathbf{Y}}$, fixed parameters $\mathbf{A}^*\boldsymbol{\beta}$, random parameters $\mathbf{A}'\mathbf{u}$ and design matrices $\mathbf{X}_{g-1}$ and $\mathbf{Z}_{g-1}$. The random error vector $\boldsymbol{\varepsilon}^*$ is normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\begin{aligned}
\text{Cov}(\boldsymbol{\varepsilon}^*) &= \text{Cov}(\mathbf{A}\bar{\boldsymbol{\varepsilon}}, \mathbf{A}\bar{\boldsymbol{\varepsilon}}) \\
&= \mathbf{A}\text{Cov}(\bar{\boldsymbol{\varepsilon}}, \bar{\boldsymbol{\varepsilon}})\mathbf{A}^T \\
&= \mathbf{A}(\frac{\sigma_\varepsilon^2}{n_0}\mathbf{I}_{gr})\mathbf{A}^T \\
&= \frac{\sigma_\varepsilon^2}{n_0}\mathbf{A}\mathbf{A}^T \\
&= \sigma_{\varepsilon^*}^2\mathbf{I}_{(g-1)r},
\end{aligned}$$

where $\sigma_{\varepsilon^*}^2 := \sigma_\varepsilon^2/n_0$. The random effects vector $\mathbf{A}'\mathbf{u}$ is normally distributed with

mean $\mathbf{0}$ and covariance matrix

$$\begin{aligned}
\mathrm{Cov}(\mathbf{A}'\mathbf{u}, \mathbf{A}'\mathbf{u}) &= (\mathbf{A}')\mathrm{Cov}(\mathbf{u}, \mathbf{u})(\mathbf{A}')^T \\
&= (\mathbf{A}')(\sigma_u^2\{(1-\rho)\mathbf{I}_{gK} + \rho(\mathbf{J}_g \otimes \mathbf{I}_K)\})(\mathbf{A}')^T \\
&= \sigma_u^2(1-\rho)(\mathbf{A}')(\mathbf{A}')^T \\
&= \sigma^2_{\mathbf{A}'\mathbf{u}}\mathbf{I}_{(g-1)K}
\end{aligned}$$

where $\sigma^2_{\mathbf{A}'\mathbf{u}} := \sigma_u^2(1-\rho)$.

Testing the (*step 2*)-hypothesis (4.3) above is equivalent to testing $\mathbf{A}'\mathbf{u} = \mathbf{0}$ which can now be tested by considering the hypothesis $H_0 : \sigma^2_{\mathbf{A}'\mathbf{u}} = 0$ in the mixed model (4.6). The test is performed again by using the Crainiceanu and Ruppert (2004) algorithm from section 4.1. The desired eigenvalues in Theorem (4.1.1) are as easily obtained as in *step 1*. This time the eigenvalues of the matrix

$$\mathrm{blockdiag}\Big\{\mathbf{C}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{C}\Big\}_{1\le i\le g}$$

need to be computed.

***Step 3.*** Given that we have failed to reject either one of the hypotheses in *step 1* or *step 2* we can model all mean curve differences with a fixed multiple regression model. This leads us to consider again the transformation $\mathbf{A}$ described in *step 2*. The mixed model (4.6) now becomes

$$\mathbf{A}\bar{\mathbf{Y}} = \mathbf{X}_{g-1}\mathbf{A}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \tag{4.7}$$

since the random effect $\mathbf{A}'\mathbf{u}$ has been declared $\mathbf{0}$. This is simply a fixed multiple regression model with response $\mathbf{A}\bar{\mathbf{Y}}$, fixed parameters $\mathbf{A}^*\boldsymbol{\beta}$ and error $\boldsymbol{\varepsilon}^*$, which is normal with mean $\mathbf{0}$ and constant variance $\sigma^2_{\varepsilon^*}$. The hypothesis (4.4) is equivalent to $H_0 : \mathbf{A}^*\boldsymbol{\beta} = \mathbf{0}$ which can be tested using methods from the vast literature on multiple regression models.

### 4.2.2 Unbalanced case

In this section we describe what happens to the three step hypotheses testing scheme in the unbalanced case; i.e. when the numbers of subjects, $n_i$ in group $i$, for $i = 1, \ldots, g$, are not necessarily equal.

*Step 1* of the algorithm is identical to *step 1* in the balanced case. The only difference lies in *step 2* and *step 3*. We consider the same model as before in (4.5)

$$\bar{\mathbf{Y}} = \mathbf{X}_g\boldsymbol{\beta} + \mathbf{Z}_g\mathbf{u} + \bar{\boldsymbol{\varepsilon}} \tag{4.8}$$

with

$$\mathbf{u} \sim \mathrm{N}\Big(\mathbf{0}, \sigma_u^2\{(1-\rho)\mathbf{I}_{gK} + \rho(\mathbf{J}_g \otimes \mathbf{I}_K)\}\Big)$$

but this time the error structure is different:

$$\bar{\boldsymbol{\varepsilon}} \sim \mathrm{N}\Big(\mathbf{0}, \mathrm{blockdiag}(\frac{\sigma_\varepsilon^2}{n_i}\mathbf{I}_r)_{1\le i\le g}\Big)$$

We apply the same transformation $\mathbf{A}$ as in the balanced case and get a similar model as in (4.6) except now we fail to achieve the iid error structure:

$$\mathbf{A}\bar{\mathbf{Y}} = \mathbf{X}_{g-1}\mathbf{A}^*\boldsymbol{\beta} + \mathbf{Z}_{g-1}\mathbf{A}'\mathbf{u} + \boldsymbol{\varepsilon}^* \tag{4.9}$$

The covariance of the random effects is as before

$$\mathrm{Cov}(\mathbf{A}'\mathbf{u}, \mathbf{A}'\mathbf{u}) = \sigma^2_{\mathbf{A}'\mathbf{u}}\mathbf{I}_{(g-1)K}$$

but the error covariance is not as simple

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\varepsilon}^*) &= \mathbf{A}(\mathrm{blockdiag}(\frac{\sigma_\varepsilon^2}{n_i}\mathbf{I}_r)_{1\le i\le g})\mathbf{A}^T \\ &= \sigma_\varepsilon^2\mathbf{S}\end{aligned}$$

where

$$\mathbf{S} = \begin{bmatrix} s_{11}\mathbf{I}_r & s_{12}\mathbf{I}_r & \dots & s_{1,g-1}\mathbf{I}_r \\ s_{21}\mathbf{I}_r & s_{22}\mathbf{I}_r & \dots & s_{2,g-1}\mathbf{I}_r \\ \vdots & \vdots & \ddots & \vdots \\ s_{g-1,1}\mathbf{I}_r & s_{g-1,2}\mathbf{I}_r & \dots & s_{g-1,g-1}\mathbf{I}_r \end{bmatrix}$$

with

$$s_{ij} = \mathbf{a}_i' \begin{pmatrix} \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_g} \end{pmatrix} \mathbf{a}_j$$

The matrix $\mathbf{S}$ is symmetric and positive definite and hence has Cholesky decomposition $\mathbf{S} = \mathbf{L}\mathbf{L}^T$. Now with a slight modification of the model (4.9) we can proceed with *steps 2* and *3* quite easily. We multiply through with $\mathbf{L}^{-1}$ to obtain:

$$\mathbf{L}^{-1}\mathbf{A}\bar{\mathbf{Y}} \;=\; (\mathbf{L}^{-1}\mathbf{X}_{g-1})\mathbf{A}^*\boldsymbol{\beta} + (\mathbf{L}^{-1}\mathbf{Z}_{g-1})\mathbf{A}'\mathbf{u} + \mathbf{L}^{-1}\boldsymbol{\varepsilon}^* \qquad (4.10)$$

which is just yet another mixed model, this time with response $\mathbf{L}^{-1}\mathbf{A}\bar{\mathbf{Y}}$, design matrices $\mathbf{L}^{-1}\mathbf{X}_{g-1}$ and $\mathbf{L}^{-1}\mathbf{Z}_{g-1}$, fixed parameters $\mathbf{A}^*\boldsymbol{\beta}$, random parameters $\mathbf{A}'\mathbf{u}$ and error vector $\mathbf{L}^{-1}\boldsymbol{\varepsilon}^*$ with the desired independence structure

$$\mathbf{L}^{-1}\boldsymbol{\varepsilon}^* \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_{(g-1)r})$$

The hypothesis in *step 2* can now be tested in the unbalanced case by testing $H_0 : \sigma^2_{\mathbf{A}'\mathbf{u}} = 0$ in the mixed model (4.10). The main difference between the tests in the balanced and in the unbalanced case is that the eigenvalues in theorem (4.1.1) are easier to calculate in the balanced case. In the unbalanced case one needs to calculate the eigenvalues of

$$\mathbf{Z}^T\mathbf{Z} - \mathbf{Z}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}$$

where $\mathbf{Z} = \mathbf{L}^{-1}\mathbf{Z}_{g-1}$ and $\mathbf{X} = \mathbf{L}^{-1}\mathbf{X}_{g-1}$. Unlike in the balanced case this matrix doesn't necessarily have any nice block-diagonal structure. This could potentially lead to computational problems when the number of groups $g$ and knots $K$ is very large.

The hypothesis in *step 3* becomes $H_0 : \mathbf{A}^*\boldsymbol{\beta} = \mathbf{0}$ in the fixed regression model

$$\mathbf{L}^{-1}\mathbf{A}\bar{\mathbf{Y}} = (\mathbf{L}^{-1}\mathbf{X}_{g-1})\mathbf{A}^*\boldsymbol{\beta} + \mathbf{L}^{-1}\boldsymbol{\varepsilon}^*$$

and this test can again be performed using standard methods from the multiple regression literature.

## 4.3 Multiple comparisons

In this section we propose a method for dealing with multiple comparisons in the functional setting. As mentioned above there is a fundamental difference between the univariate and the functional setting since in the latter case one is dealing with a continuum of means rather than a single mean per group. The method proposed here involves constructing simultaneous confidence bands. The construction is an extension of the construction of simultaneous confidence bands for single curves discussed in Ruppert et al. (2003).

As an example consider a setting where there are 4 groups and the experimenter is interested in the following comparisons,

$$\begin{aligned} f_1 &= f_2, \\ f_1 &= f_3, \\ f_1 &= f_4. \end{aligned}$$

The idea behind the method, discussed in detail below, is to simulate simultaneous confidence bands for $f_1 - f_2$, $f_1 - f_3$ and $f_1 - f_4$. Then the experimenter can see

which parts of these mean difference bands contain 0. Based on these simultaneous confidence bands one can make a statement about all the mean differences at any parts of the curves with $100(1-\alpha)\%$ confidence.

Consider the more general framework where the comparison of interest is formed with a contrast matrix

$$\mathbf{L} = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_c' \end{bmatrix}$$

where $c$ is the number of comparisons and $\mathbf{a}_k = (a_{k1}, \ldots, a_{kg})'$, for $k = 1, \ldots, c$. The comparison of interest can now be written as

$$\mathbf{L}f = \mathbf{0}$$

where $f = (f_1, \ldots, f_g)'$.

We want to simulate simultaneous confidence bands for $\mathbf{L}f$ over a grid of $M$ $x$-values

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)'$$

The representation here is a fairly natural generalization of the representation in Ruppert et al. (2003) of a simultaneous confidence band for a single curve. For $i = 1, \ldots, g$ let

$$\mathbf{f}_{i\boldsymbol{\xi}} = \begin{bmatrix} f_i(\xi_1) \\ \vdots \\ f_i(\xi_M) \end{bmatrix}$$

be the true mean function for group $i$ over $\boldsymbol{\xi}$. Let $\hat{\mathbf{f}}_{i\boldsymbol{\xi}}$ be the corresponding EBLUP for group $i$ over $\boldsymbol{\xi}$. We then have

$$\hat{\mathbf{f}}_{i\boldsymbol{\xi}} - \mathbf{f}_{i\boldsymbol{\xi}} = \mathbf{C}_{\boldsymbol{\xi}} \begin{bmatrix} \hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i \\ \hat{\mathbf{u}}_i - \mathbf{u}_i \end{bmatrix}$$

for all $i = 1, \ldots, g$, where

$$\mathbf{C}_{\boldsymbol{\xi}} = \begin{bmatrix} \mathbf{1} & \boldsymbol{\xi} & \boldsymbol{\xi}^2 & (\boldsymbol{\xi} - \tau_1\mathbf{1})_+^2 & \ldots & (\boldsymbol{\xi} - \tau_K\mathbf{1})_+^2 \end{bmatrix}.$$

Now define $\mathbf{f}_{\boldsymbol{\xi}} = (\mathbf{f}_{1\boldsymbol{\xi}}, \ldots, \mathbf{f}_{g\boldsymbol{\xi}})'$, and similarly define $\hat{\mathbf{f}}_{\boldsymbol{\xi}}$. Then we can write more compactly,

$$\hat{\mathbf{f}}_{\boldsymbol{\xi}} - \mathbf{f}_{\boldsymbol{\xi}} = \mathbf{C}^*_{\boldsymbol{\xi}} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\mathbf{u}}_1 - \mathbf{u}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g \\ \hat{\mathbf{u}}_g - \mathbf{u}_g \end{bmatrix}, \tag{4.11}$$

where $\mathbf{C}^*_{\boldsymbol{\xi}} = \text{blockdiag}(\mathbf{C}_{\boldsymbol{\xi}})_{1 \le j \le g}$.

Define $\mathbf{S} = [\mathbf{B} \quad \mathbf{C}]$ and let

$$\mathbf{S}^* = \text{blockdiag}(\mathbf{1}_{n_i} \otimes \mathbf{S})_{1 \le i \le g}$$

and

$$\mathbf{D}^* = \text{blockdiag}(\mathbf{D})_{1 \le i \le g}$$

where $\mathbf{D} = \text{diag}(0, 0, 0, 1, \ldots, 1)$. The three zeros correspond to the fixed parameters and the $K$ ones correspond to the random parameters. It is easily verified that

$$\left((\mathbf{S}^*)^T\mathbf{S}^* + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\mathbf{D}^*\right)^{-1} = \text{blockdiag}\left\{\frac{1}{n_i}\left(\mathbf{S}^T\mathbf{S} + \frac{\sigma_\varepsilon^2/n_i}{\sigma_u^2}\mathbf{D}\right)^{-1}\right\}.$$

Defining

$$\mathbf{v}_{\boldsymbol{\beta}\mathbf{u}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\mathbf{u}}_1 - \mathbf{u}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g \\ \hat{\mathbf{u}}_g - \mathbf{u}_g \end{bmatrix}$$

we recall from (4.11):

$$\hat{\mathbf{f}}_{\boldsymbol{\xi}} - \mathbf{f}_{\boldsymbol{\xi}} = \mathbf{C}^*_{\boldsymbol{\xi}} \mathbf{v}_{\boldsymbol{\beta}\mathbf{u}},$$

and note that

$$\mathbf{v}_{\boldsymbol{\beta}\mathbf{u}} \overset{approx.}{\sim} \mathbf{N}\Big(\mathbf{0}, \text{blockdiag}\left\{\frac{\hat{\sigma}^2_\varepsilon}{n_i}\left(\mathbf{S}^T\mathbf{S} + \frac{\hat{\sigma}^2_\varepsilon/n_i}{\hat{\sigma}^2_u}\mathbf{D}\right)^{-1}\right\}\Big). \tag{4.12}$$

which was to be expected since the EBLUPs $\hat{\boldsymbol{\beta}}_i$ and $\hat{\mathbf{u}}_i$ from (3.15), for a given $\alpha$, are equal to the EBLUPs we obtain from the mean profile mixed model discussed in section 3.4.

Now define $\mathbf{L}^* = \mathbf{L} \otimes \mathbf{I}_M$. Then evidently a $100(1-\alpha)\%$ simultaneous confidence band for $\mathbf{L}^*\mathbf{f}_{\boldsymbol{\xi}}$ is

$$\mathbf{L}^*\hat{\mathbf{f}}_{\boldsymbol{\xi}} \pm m_{1-\alpha}\Big(\widehat{\mathbf{st.dev.}}\{\mathbf{L}^*[\hat{\mathbf{f}}_{\boldsymbol{\xi}} - \mathbf{f}_{\boldsymbol{\xi}}]\}\Big)$$

where $\widehat{\mathbf{st.dev.}}\{\mathbf{L}^*[\hat{\mathbf{f}}_{\boldsymbol{\xi}} - \mathbf{f}_{\boldsymbol{\xi}}]\}$ are the square roots of the diagonal elements of the $cM \otimes cM$ matrix

$$\begin{aligned} & (\mathbf{L}^*\mathbf{C}^*_{\boldsymbol{\xi}})\text{blockdiag}\left\{\frac{\hat{\sigma}^2_\varepsilon}{n_i}\Big(\mathbf{S}^T\mathbf{S} + \frac{\hat{\sigma}^2_\varepsilon/n_i}{\hat{\sigma}^2_u}\mathbf{D}\Big)^{-1}\right\}(\mathbf{L}^*\mathbf{C}^*_{\boldsymbol{\xi}})^T \\ = \quad & (\mathbf{L}^*)\text{blockdiag}\left\{\frac{\hat{\sigma}^2_\varepsilon}{n_i}\mathbf{C}_{\boldsymbol{\xi}}\Big(\mathbf{S}^T\mathbf{S} + \frac{\hat{\sigma}^2_\varepsilon/n_i}{\hat{\sigma}^2_u}\mathbf{D}\Big)^{-1}\mathbf{C}^T_{\boldsymbol{\xi}}\right\}(\mathbf{L}^*)^T \end{aligned}$$

and $m_{1-\alpha}$ is the $(1-\alpha)$ quantile of the random variable

$$\max_j \left( \sup_{x \in \mathcal{X}} \left| \frac{\mathbf{a}_j'[\hat{f}(x) - f(x)]}{\widehat{\text{st.dev.}}\{\mathbf{a}_j'[\hat{f}(x) - f(x)]\}} \right| \right),$$

which can be approximated by

$$\max_{1 \le l \le cM} \left| \frac{\left( \mathbf{L}^* \mathbf{C}^*_{\boldsymbol{\xi}} \mathbf{v}_{\boldsymbol{\beta}_\mathbf{u}} \right)_l}{\left( \widehat{\mathbf{st.dev.}}\{\mathbf{L}^*[\hat{\mathbf{f}}_{\boldsymbol{\xi}} - \mathbf{f}_{\boldsymbol{\xi}}]\} \right)_l} \right|. \tag{4.13}$$

One can now simulate a realization of (4.12) a large number of times; Ruppert et al. (2003) suggest $N = 10,000$. This involves simulating independently

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i \\ \hat{\mathbf{u}}_i - \mathbf{u}_i \end{bmatrix} \overset{approx.}{\sim} \mathbf{N} \left\{ \mathbf{0}, \frac{\hat{\sigma}^2_\varepsilon}{n_i} \left( \mathbf{S}^T \mathbf{S} + \frac{\hat{\sigma}^2_\varepsilon / n_i}{\hat{\sigma}^2_u} \mathbf{D} \right)^{-1} \right\},$$

for $i = 1, \ldots, g$. For each realization one computes the value of (4.13). Then at the end of the simulation the values are sorted and the $(1 - \alpha)$ sample quantile is used as $m_{1-\alpha}$.

# Chapter 5

# Case study

In this chapter we will apply the methods, described in chapter 3 and 4 on a functional data set from Ramsay and Silverman (1997). The data consists of average daily temperatures, measured over a one year time period, at weather stations distributed across Canada. The country is divided into $g = 4$ geographical regions, Atlantic, Continental, Pacific and Arctic region. The regions have $n_1 = 15$, $n_2 = 12$, $n_3 = 5$ and $n_4 = 3$ weather stations respectively, total of $N = \sum_{i=1}^{4} n_i = 35$ stations. At each weather station, average daily temperature is measured over the $r = 365$ days of the year. We can see one such temperature profile, for a weather station in the Atlantic region, in figure 5.1. When we stack all 35 temperature profiles together, we end up with a response vector, $\mathbf{Y}$, of length, $n = Nr = 12775$.

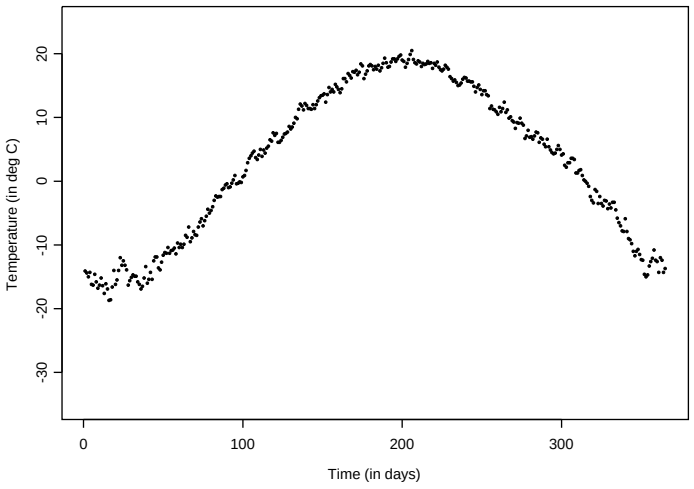


Figure 5.1: A temperature profile, of a single weather station, in the Atlantic region of Canada. Temperatures are measured in deg C over a one year period.

## 5.1 Fitting the model to the Canadian weather data

Let $\mathbf{y}_{ij}(t_k)$ denote the average temperature measured at day, $k = 1, \ldots, 365$, in weather station, $j = 1, \ldots, n_i$, within region, $i = 1, \ldots, 4$. As discussed in section 2.3, we wish to model our data according to the mixed model

$$\mathbf{y}_{ij}(t_k) = \beta_{0i} + \beta_{1i} t_k + \beta_{2i} t_k^2 + \sum_{m=1}^{K} u_{mi}(t_k - \tau_m)_+^2 + \varepsilon_{ijk},$$

For the Canadian weather data, we place $K = 11$ knots, $\tau_m$, evenly across the time interval, $[1, 365]$, with each knot, roughly located, between the end and the beginning of two consecutive months.

We fit the model above, with the REML criteria, using the Newton-Raphson algorithm to maximize the restricted log-likelihood. The necessary computing formulas are given in section (3.3). The REML estimates of the variance components and the value of the restricted log-likelihood, are given in table 5.1. We can use these REML estimates directly to obtain the EBLUP estimates, (3.15). The EBLUPs determine the fitted mean temperature functions for each of the four regions and the resulting fits are plotted in figure 5.2, along with plots of the raw discrete data. As we can see, the fits are rather pleasing and seem to capture well the regional weather trends of the underlying data.

Table 5.1: The resulting REML estimates of the variance components for the Canadian weather data and the value of the restricted log-likelihood evaluated at those estimates.

| | |
|---|---|
| $\hat{\sigma}_\varepsilon^2$ | 14.44 |
| $\hat{\sigma}_u^2$ | $3.96 \cdot 10^{-6}$ |
| $l_R(\hat{\sigma}_\varepsilon^2, \hat{\sigma}_u^2)$ | $-35,267.37$ |

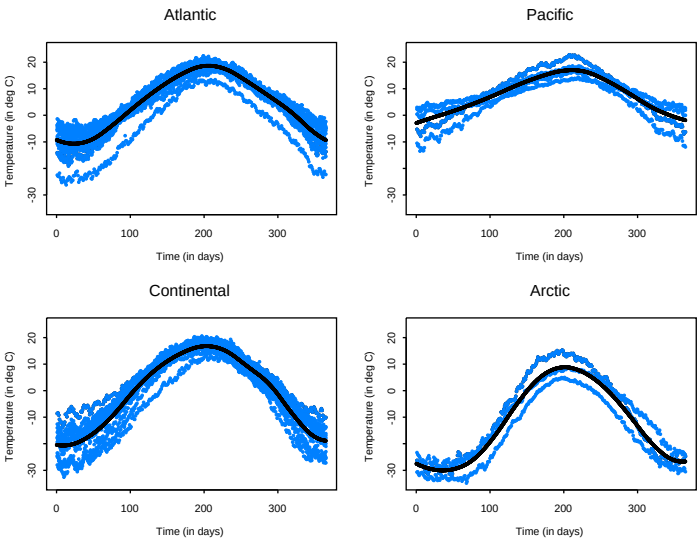


Figure 5.2: The fitted mean temperature curves, for each weather region, Atlantic, Continental, Pacific and Arctic region. The raw data is also plotted, to asses the goodness of the fit, graphically.

## 5.2 Hypotheses testing

In this section we will address the problem of testing the hypothesis

$$H_0 : f_1 = \ldots = f_4,$$

for the Canadian weather data, using the hypotheses testing scheme proposed in chapter 4. In essence what the hypothesis above implies is that the average daily temperature behaves in the same manner across the four geographic regions in Canada.

***Step 1.*** We start with *step 1.* and test hypothesis (4.2)

$$H_0 : \mathbf{u}_1 = \ldots = \mathbf{u}_4 = 0,$$

or equivalently

$$H_0 : \sigma_u^2 = 0.$$

This hypothesis implies that there is no smoothing to be done and a quadratic model is sufficient to model the weather data. Looking at the graphs in figure 5.2, we expect to reject this hypothesis, beforehand.

To use the results of Crainiceanu and Ruppert (2004), in theorem 4.1.1, we need to calculate the eigenvalues of the $44 \times 44$-matrix

$$\text{blockdiag}\Big\{n_i \mathbf{C}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{C}\Big\}_{1\leq i\leq 4}$$

which is equivalent to calculating the eigenvalues of the $11 \times 11$-matrices

$$n_i \mathbf{C}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{C}$$

for $n_1 = 15$, $n_2 = 12$, $n_3 = 5$ and $n_4 = 3$, separately. Once the eigenvalues are obtained, one can simulate the null finite sampling distribution of the restricted likelihood ratio test statistic (RLRT), using the 6 step algorithm, described in section 4.1. For step 1 of the simulation algorithm, as suggested in Crainiceanu and Ruppert (2004), we used for $\lambda_i > 0$, 200 grid points, equally spaced on the natural log-scale between $[-12, 12]$. All the other steps of the algorithm are easily implemented. The number of realizations was $10,000$. A histogram of the $10,000$ realizations can be seen in figure 5.3. There is a point mass at 0, with relative frequency 0.6104.

To construct the restricted likelihood ratio statistic, we need the null restricted log-likelihood. When $\sigma_u^2 = 0$, the model, discussed in this thesis, becomes a standard linear regression model. In terms of our notation the following holds, $\mathbf{\Psi}(\alpha) = \mathbf{I}$, and the formulas for the fixed parameters simplify to

$$\boldsymbol{\beta}_i(\alpha) = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\bar{\mathbf{y}}_{i\cdot},$$

for $i = 1, \ldots, 4$. Hence the restricted log-likelihood from (3.1) simplifies to

$$l_R^0(\sigma_\varepsilon^2) = -\frac{1}{2}[(n-p)\log(\sigma_\varepsilon^2) + \sum_{i=1}^{4} \mathbf{S}_{i0}(\mathbf{y}, \sigma_\varepsilon^2)/\sigma_\varepsilon^2 + \sum_{i=1}^{4} \log|n_i \mathbf{B}^T\mathbf{B}|] - \frac{n}{2}\log(2\pi),$$

where

$$\begin{aligned}\mathbf{S}_{i0}(\mathbf{y}, \sigma_\varepsilon^2) &= \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}) \\ &\quad + n_i(\bar{\mathbf{y}}_{i.} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\bar{\mathbf{y}}_{i.})^T(\bar{\mathbf{y}}_{i.} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\bar{\mathbf{y}}_{i.}).\end{aligned}$$

The maximum of the null restricted log-likelihood above is $-38,938.77$, maximized at $\hat{\sigma}_{\varepsilon,0}^2 = 25.94$. This results in the test statistic

$$\begin{aligned}RLRT &= -2(l_R^0(\hat{\sigma}_{\varepsilon,0}^2) - l_R(\hat{\alpha})) \\ &= -2\left(-38,938.77 - (-35,267.37)\right) \\ &= 7342.80.\end{aligned}$$

Comparing this test statistic to the finite sample distribution of RLRT, (see figure 5.3), we reject the null hypothesis of no smoothing, with a p-value of 0.

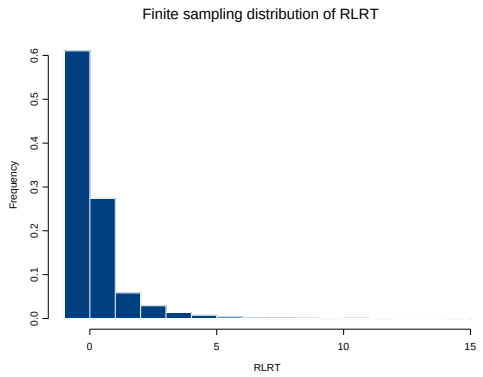


Figure 5.3: A histogram of the simulated finite sample distribution of the restricted likelihood ratio statistic, for testing the *step 1*-hypothesis. There is a point mass at zero with relative frequency 0.6104.

***Step 2.*** Since we rejected the *step 1* hypothesis, we conclude that a quadratic model does not fit the data well, and move on to *step 2*. We test the hypothesis

$$H_0 : \mathbf{u}_1 = \ldots = \mathbf{u}_g,$$

which implies that all mean differences can be modeled with a quadratic model.

We consider the transformed mean model (4.9)

$$\mathbf{A}\bar{\mathbf{Y}} = \mathbf{X}_{g-1}\mathbf{A}^*\boldsymbol{\beta} + \mathbf{Z}_{g-1}\mathbf{A}'\mathbf{u} + \boldsymbol{\varepsilon}^*,$$

where

$$\mathbf{A}\bar{\mathbf{Y}} = \begin{bmatrix} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_4)/\sqrt{2} \\ (\bar{\mathbf{Y}}_2 - \bar{\mathbf{Y}}_3)/\sqrt{2} \\ (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 - \bar{\mathbf{Y}}_3 + \bar{\mathbf{Y}}_4)/2 \end{bmatrix}, \quad \boldsymbol{\varepsilon}^* = \begin{bmatrix} (\bar{\boldsymbol{\varepsilon}}_1 - \bar{\boldsymbol{\varepsilon}}_4)/\sqrt{2} \\ (\bar{\boldsymbol{\varepsilon}}_2 - \bar{\boldsymbol{\varepsilon}}_3)/\sqrt{2} \\ (\bar{\boldsymbol{\varepsilon}}_1 - \bar{\boldsymbol{\varepsilon}}_2 - \bar{\boldsymbol{\varepsilon}}_3 + \bar{\boldsymbol{\varepsilon}}_4)/2 \end{bmatrix}$$

and

$$\mathbf{A}^*\boldsymbol{\beta} = \begin{bmatrix} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_4)/\sqrt{2} \\ (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_3)/\sqrt{2} \\ (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 - \boldsymbol{\beta}_3 + \boldsymbol{\beta}_4)/2 \end{bmatrix}, \quad \mathbf{A}'\mathbf{u} = \begin{bmatrix} (\mathbf{u}_1 - \mathbf{u}_4)/\sqrt{2} \\ (\mathbf{u}_2 - \mathbf{u}_3)/\sqrt{2} \\ (\mathbf{u}_1 - \mathbf{u}_2 - \mathbf{u}_3 + \mathbf{u}_4)/2 \end{bmatrix}.$$

The random effects, $\mathbf{A}'\mathbf{u}$, are normally distributed with mean $\mathbf{0}$ and covariance

$$\mathrm{Cov}(\mathbf{A}'\mathbf{u}, \mathbf{A}'\mathbf{u}) = \sigma^2_{\mathbf{A}'\mathbf{u}}\mathbf{I}_{3K}.$$

The random vector, $\boldsymbol{\varepsilon}^*$, is normally distributed with mean 0 and covariance matrix

$$\mathrm{Cov}(\boldsymbol{\varepsilon}^*) = \sigma^2_{\varepsilon}\mathbf{S}$$

with

$$\mathbf{S} = \begin{bmatrix} \frac{1}{2}(\frac{1}{n_1} + \frac{1}{n_4})\mathbf{I}_r & \mathbf{0} & \frac{1}{2\sqrt{2}}(\frac{1}{n_1} - \frac{1}{n_4})\mathbf{I}_r \\ \mathbf{0} & \frac{1}{2}(\frac{1}{n_2} + \frac{1}{n_3})\mathbf{I}_r & \frac{1}{2\sqrt{2}}(\frac{1}{n_3} - \frac{1}{n_2})\mathbf{I}_r \\ \frac{1}{2\sqrt{2}}(\frac{1}{n_1} - \frac{1}{n_4})\mathbf{I}_r & \frac{1}{2\sqrt{2}}(\frac{1}{n_3} - \frac{1}{n_2})\mathbf{I}_r & \frac{1}{4}(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4})\mathbf{I}_r \end{bmatrix}.$$

We find the Cholesky decomposition of $\mathbf{S} = \mathbf{L}\mathbf{L}^T$ and form the mixed model (4.10)

$$\mathbf{L}^{-1}\mathbf{A}\bar{\mathbf{Y}} = (\mathbf{L}^{-1}\mathbf{X}_{g-1})\mathbf{A}^*\boldsymbol{\beta} + (\mathbf{L}^{-1}\mathbf{Z}_{g-1})\mathbf{A}'\mathbf{u} + \mathbf{L}^{-1}\boldsymbol{\varepsilon}^*, \tag{5.1}$$

with

$$\mathbf{L}^{-1}\boldsymbol{\varepsilon}^* \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_{3r}).$$

The *step 2* hypothesis can now be tested, using the simulation algorithm, by testing $H_0 : \sigma^2_{\mathbf{A}'\mathbf{u}} = 0$ in the mixed model above. We again use $10,000$ realizations and obtain the simulated finite sample distribution of the RLRT, see figure 5.4. We obtain the test statistic

$$\begin{aligned} RLRT &= -2(l_R^0(\hat{\sigma}^2_{\varepsilon,0}) - l_R(\hat{\alpha})) \\ &= -2\left(-3,109.17 - (-2,094.44)\right) \\ &= 2,029.46 \end{aligned}$$

and based on this statistic, we reject the null hypothesis with a p-value of 0. This result agrees with what we observe in figure 5.5. Based on that plot, it is clear that there is considerable smoothing involved, when fitting the model (5.1).

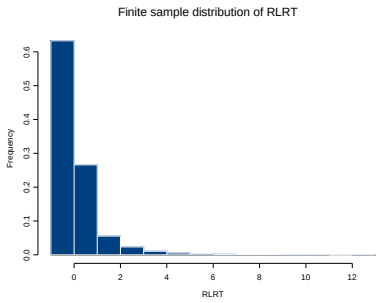


Figure 5.4: A histogram of the simulated finite sample distribution of the restricted likelihood ratio statistic, for testing the *step 2*-hypothesis. There is a point mass at zero with relative frequency 0.6326.

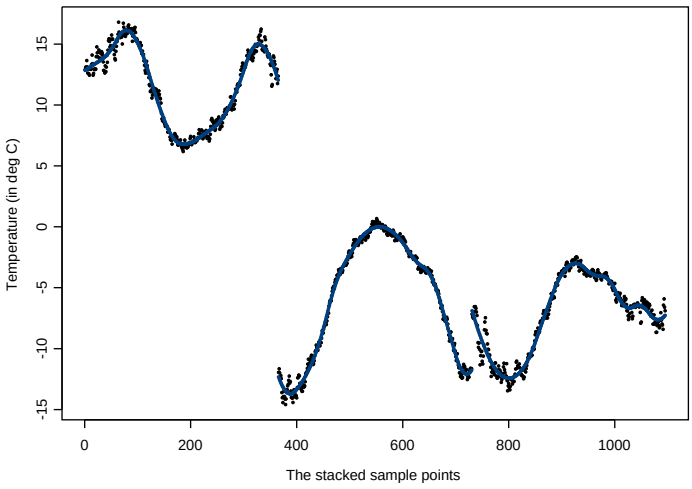


Figure 5.5: A plot of the transformed mean vector, $\mathbf{A}\bar{\mathbf{Y}}$, against the stacked sample points. The three curves correspond to $(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_4)/\sqrt{2}$, $(\bar{\mathbf{Y}}_2 - \bar{\mathbf{Y}}_3)/\sqrt{2}$ and $(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 - \bar{\mathbf{Y}}_3 + \bar{\mathbf{Y}}_4)/2$, respectively. The dots represent the raw transformed means, and the solid curves correspond to the fitted values from the model, (5.1).

## 5.3 Multiple comparisons

Since we rejected the hypothesis in *step 2*, it is clear that the mean temperature profiles are different, across weather regions, and thus no need to proceed with *step 3*. Instead, we consider multiple comparisons, to see where the differences lie. As discussed in section 4.3, one way to do multiple comparisons is to look at simulated simultaneous confidence bands about linear contrasts of the mean curves. In this section we will consider the following multiple comparison

$$f_3 = f_1,$$

$$f_3 = f_2,$$

$$f_3 = f_4.$$

This comparison is, in essence, comparing the average temperature in the Pacific region to the ones of the other three geographic regions. We construct the contrast

$$\mathbf{L} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

and then use the simulation method described in section 4.3. We simulate over a grid of $x$-values, simultaneous 95% confidence bands for the linear contrasts, $f_3 - f_1$, $f_3 - f_2$ and $f_3 - f_4$. The sample points, $1, \ldots, 365$, were used as grid points in the simulation and the number of realizations was $10,000$. The results can be seen in figure 5.6. We can now make statements, with 95% confidence, about the three mean profile differences at any time point. We can see that the temperature is higher in the Pacific region, than in the Arctic region, all year round. In the winter, the temperature is also higher in the Pacific region, than in both the Atlantic and the Continental region. In the summer, on the other hand, the Atlantic and the Continental region catch up. The difference between the Pacific and the Continental temperatures, over the summer months, is not statistically significant and the temperature in the Atlantic region even exceeds the Pacific temperature, over a roughly month period, sometime around July.

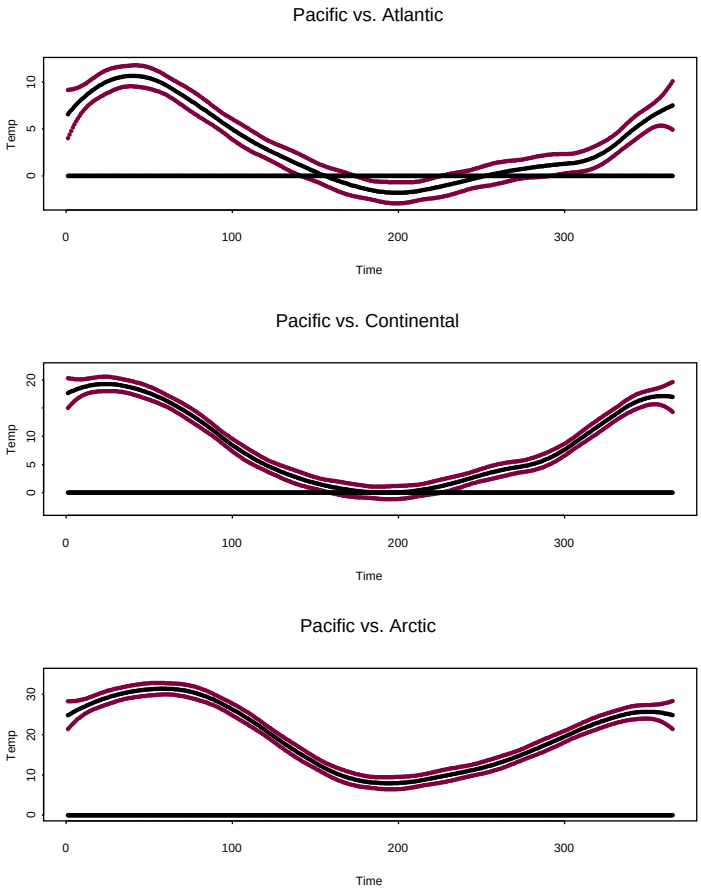


Figure 5.6: The three plots show the fitted mean curve differences, $\hat{f}_3 - \hat{f}_1$, $\hat{f}_3 - \hat{f}_2$ and $\hat{f}_3 - \hat{f}_4$ respectively, along with simulated simultaneous 95% confidence bands.

# Chapter 6

# Discussion

In this thesis we extended the idea of using penalized splines as BLUPs for smoothing individual curves. We presented a way of smoothing several mean curves across different groups using a single mixed model. We showed how the large dimensionality of the estimation problem could be reduced substantially, making the problem computationally tractable. We then showed that if the number of curves is large, the mean model may be used in place of the full model thus simplifying the problem substantially. Fitting the mean model will allow the researcher to make use of standard softwares such as SAS or S-plus. For fitting the mean profile model of section 3.4, for example, one could use the lme function of S-plus. Using the proposed mixed model, we constructed a hypotheses testing scheme for the problem of testing equality of mean curves. In the process, we established a connection between the FANOVA problem and the recent theory of Crainiceanu and Ruppert (2004) for testing variance components of a mixed model being zero. We then proposed a way of doing multiple comparisons in the functional setting, again using the underlying mixed model. The idea involved the simulation of simultaneous confidence bands around mean curve contrasts. This is a natural way of comparing quantitative differences between curves, but might not work well if the differences are of other nature. Finally we used the methods described in this thesis to analyze the Canadian Weather data of Ramsay and Silverman (1997). This is a typical functional data set where the sample points, $(t_k)_{k=1}^r$, are the same across all curves.

Although it is often the case that functional responses of a certain data set are defined on the same set of design points, this is not always the case. However,

it should be a straightforward task to extend the ideas presented in this thesis to functional data sets, where the responses are defined at different sample points. The problem of such an extension though, is that we will lose the nice structure of our design matrices, and hence, the estimation procedure becomes more computationally expensive. The setup of Brumback and Rice (1998) allowed for different design points between curves and they proposed a very efficient EM-algorithm for doing the REML estimations. Since the penalized splines are computationally more desirable than the smoothing splines, it should be easy to adapt their algorithm to the penalized splines setup.

We mentioned in section 2.3 the problem with the independence assumption in the functional mixed model, (2.12). In practice it may not be the case that $\varepsilon_{ijk}$ is independent of $\varepsilon_{ijk'}$. A possible remedy is to assume that the error vectors $\boldsymbol{\varepsilon}_{ij}$ are identical normally distributed with mean $\mathbf{0}$ and covariance matrix $\sigma_\varepsilon^2\Sigma$. One could then estimate the correlation matrix either using some time-series methods, or just by calculating sample correlation. Our model would take on the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

with

$$\boldsymbol{\varepsilon} \sim \mathbf{N}\left(\mathbf{0}, \sigma_\varepsilon^2 \text{blockdiag}(\boldsymbol{\Sigma})\right) \quad \text{and} \quad \mathbf{u} \sim \mathbf{N}\left(\mathbf{0}, \mathbf{I}_{gK}\right).$$

We would then obtain the Cholesky decomposition of $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$ and multiply through the mixed model with blockdiag($\mathbf{L}^{-1}$). This would ensure the same block-diagonal structure of the design matrices and result in the desired diagonal error covariance matrix.

The problem of testing the hypothesis of two curves being equal, might seem conceptually well defined: either the curves are equal, or they are not. However,

there are some subtleties that require further attention. What does it really mean when we say two curves are different? As discussed in the beginning of chapter 4, the curves could agree on some parts, but deviate from each other on other parts. Another phenomena that we mentioned in the introduction, is that it is possible to declare two curves different, but still see no significant differences anywhere on the curves when considering simultaneous confidence bands about the mean curve differences. An example where such a phenomena could occur is shown in figure 6.1. The researcher, in this setting would probably be more interested in looking at derivatives when trying to detect how the curves are different. This example demonstrates that care must be taken when analyzing differences between curves. Ultimately, the FANOVA problem is by no means straightforward, and there is a need for a more solid framework to resolve it. It will be interesting to see how this problem will be addressed in the future as the field of functional data analysis progresses.

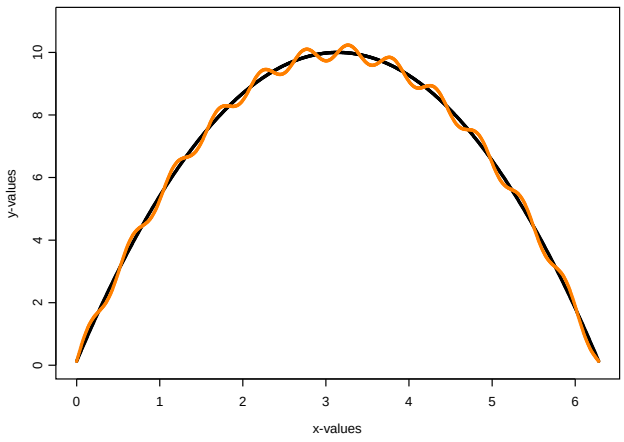


Figure 6.1: An example, where the curves are clearly different, but their point-wise differences are probably not statistically significant at any $x$-value.

# Appendix A

In section 3.3 simplified form of the Restricted log-likelihood was stated along with its first and second derivatives. In this appendix we will derive in detail all the formulas of section 3.3, (3.9)-(3.15) that were stated without justification. Before continuing it might be useful to state three basic rules of matrix differentiation used throughout. Let $\mathbf{A}(\alpha)$ be an invertible matrix which has elements that depend on the variable $\alpha$. Then

$$\frac{d}{d\alpha}\mathbf{A}^{-1}(\alpha) = -\mathbf{A}^{-1}\frac{d\mathbf{A}}{d\alpha}\mathbf{A}^{-1} \tag{A.1}$$

$$\frac{d}{d\alpha}\log|\mathbf{A}(\alpha)| = \mathrm{tr}\{\mathbf{A}^{-1}\frac{d\mathbf{A}}{d\alpha}\} \tag{A.2}$$

Let $\mathbf{B}(\alpha)$ be another matrix depending on $\alpha$ and let us assume that the dimensions of the matrices $\mathbf{A}(\alpha)$ and $\mathbf{B}(\alpha)$ match. Then the chain rule for matrix differentiation is:

$$\frac{d}{d\alpha}\Big(\mathbf{A}(\alpha)\mathbf{B}(\alpha)\Big) = \Big(\frac{d}{d\alpha}\mathbf{A}(\alpha)\Big)\mathbf{B}(\alpha) + \mathbf{A}(\alpha)\Big(\frac{d}{d\alpha}\mathbf{B}(\alpha)\Big) \tag{A.3}$$

The only thing that a practitioner has to worry about when dealing with matrix derivatives is to make sure that the order of terms is maintained since matrix multiplication is not commutative.

## A.1

In this section we will consider the derivation of $\mathbf{A}_{1i}$, $\mathbf{W}_i$ and $\mathbf{A}_{2i}$ and their first and second derivatives, (3.9)-(3.11). Before doing so let us first derive the form of $\mathbf{\Psi}_i(\alpha)$, recall

$$\mathbf{A}_i(\alpha) = \alpha\mathbf{I}_K, \quad \mathbf{X}_i = \mathbf{1}_{n_i} \otimes \mathbf{B} \quad \text{and} \quad \mathbf{Z}_i = \mathbf{1}_{n_i} \otimes \mathbf{C}$$

then

$$\begin{aligned}
\boldsymbol{\Psi}_i(\alpha) &= \mathbf{I} - \mathbf{Z}_i\{\mathbf{A}_i(\alpha) + \mathbf{Z}_i^T\mathbf{Z}_i\}^{-1}\mathbf{Z}_i^T \\
&= \mathbf{I} - (\mathbf{1}_{n_i} \otimes \mathbf{C})\{\alpha\mathbf{I} + n_i\mathbf{C}^T\mathbf{C}\}^{-1}(\mathbf{1}_{n_i}^T \otimes \mathbf{C}^T) \\
&= \mathbf{I} - \mathbf{J}_{n_i} \otimes \left(\mathbf{C}\{\alpha\mathbf{I} + n_i\mathbf{C}^T\mathbf{C}\}^{-1}\mathbf{C}^T\right) \\
&= \mathbf{I} - \mathbf{J}_{n_i} \otimes \mathbf{A}_{1i} \qquad \text{(A.4)}
\end{aligned}$$

(3.9): The derivatives of the matrix $\mathbf{A}_{1i} = \frac{1}{n_i}\mathbf{C}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T$ follow trivially from matrix rule $(A.1)$ and the chain rule $(A.3)$:

$$\begin{aligned}
\mathbf{A}_{1i} &= \frac{1}{n_i}\mathbf{C}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T \\
d\mathbf{A}_{1i} &= \frac{d}{d\alpha}\mathbf{A}_{1i} \\
&= -\frac{1}{n_i^2}\mathbf{C}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T \\
d^2\mathbf{A}_{1i} &= \frac{d^2}{d\alpha^2}\mathbf{A}_{1i} \\
&= \frac{2}{n_i^3}\mathbf{C}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}\mathbf{C}^T \qquad \text{(A.5)}
\end{aligned}$$

(3.10): The derivatives of the matrix $\mathbf{W}_i = \mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i$ also follow trivially from matrix rule $(A.1)$ and the chain rule $(A.3)$ once the form of $\mathbf{W}_i$ has been derived:

$$\begin{aligned}
\mathbf{W}_i &= \mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i \\
&= (\mathbf{1}_{n_i}^T \otimes \mathbf{B}^T)[\mathbf{I} - \mathbf{J}_{n_i} \otimes \mathbf{A}_{1i}](\mathbf{1}_{n_i} \otimes \mathbf{B}) \\
&= n_i\mathbf{B}^T\mathbf{B} - n_i^2\mathbf{B}^T\mathbf{A}_{1i}\mathbf{B} \\
&= n_i\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i})\mathbf{B} \\
d\mathbf{W}_i &= \frac{d}{d\alpha}\mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i \\
&= -n_i^2\mathbf{B}^T d\mathbf{A}_{1i}\mathbf{B}
\end{aligned}$$

$$
\begin{aligned}
d^2\mathbf{W}_i &= \frac{d^2}{d\alpha^2}\mathbf{X}_i^T\mathbf{\Psi}_i(\alpha)\mathbf{X}_i \\
&= -n_i^2\mathbf{B}^T d^2\mathbf{A}_{1i}\mathbf{B} \qquad \text{(A.6)}
\end{aligned}
$$

(3.11): The derivatives of the matrix $\mathbf{A}_{2i} = \mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i})$ are a little less trivial but follow just as before from matrix rule $(A.1)$ and the chain rule $(A.3)$:

$$
\begin{aligned}
\mathbf{A}_{2i} &= \mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i}) \\
d\mathbf{A}_{2i} &= \frac{d}{d\alpha}\mathbf{A}_{2i} \\
&= -\mathbf{B}(\mathbf{W}_i)^{-1}(d\mathbf{W}_i)(\mathbf{W}_i)^{-1}\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i}) \\
&\quad -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i} \\
&= -\mathbf{B}(\mathbf{W}_i)^{-1}(-n_i^2\mathbf{B}^T d\mathbf{A}_{1i}\mathbf{B})(\mathbf{W}_i)^{-1}\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i}) \\
&\quad -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i} \\
&= -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i} \\
&\quad n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i}\Big\{n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T(\mathbf{I} - n_i\mathbf{A}_{1i})\Big\} \\
&= -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i}) \\
d^2\mathbf{A}_{2i} &= \frac{d^2}{d\alpha^2}\mathbf{A}_{2i} \\
&= n_i\mathbf{B}(\mathbf{W}_i)^{-1}(d\mathbf{W}_i)(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i}) \\
&\quad -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d^2\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i}) \\
&\quad -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i}(-n_i d\mathbf{A}_{2i}) \\
&= n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T(n_i d\mathbf{A}_{1i})\{-n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i})\} \\
&\quad -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d^2\mathbf{A}_{1i}(\mathbf{I} - n_i\mathbf{A}_{2i}) \\
&\quad +n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T(n_i d\mathbf{A}_{1i})d\mathbf{A}_{2i}
\end{aligned}
$$

$$
\begin{aligned}
&= n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T(n_i d\mathbf{A}_{1i})\{d\mathbf{A}_{2i}\} \\
&\quad -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T d^2\mathbf{A}_{1i}(\mathbf{I}-n_i\mathbf{A}_{2i}) \\
&\quad +n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T(n_i d\mathbf{A}_{1i})d\mathbf{A}_{2i} \\
&= -n_i\mathbf{B}(\mathbf{W}_i)^{-1}\mathbf{B}^T\Big\{d^2\mathbf{A}_{1i}(\mathbf{I}-n_i\mathbf{A}_{2i}) - 2n_i d\mathbf{A}_{1i}d\mathbf{A}_{2i}\Big\} \qquad \text{(A.7)}
\end{aligned}
$$

## A.2

In this section the computation formulas for the restricted log-likelihood (3.12), and the first and second derivatives (3.13) and (3.14) will be derived from the already established log-likelihood (3.6).

(3.15): But first let us derive the EBLUP formulas for $\boldsymbol{\beta}_i$ and $\mathbf{u}_i$:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_i &= \{\mathbf{X}_i^T\boldsymbol{\Psi}_i(\hat{\alpha})\mathbf{X}_i\}^{-1}\mathbf{X}_i^T\boldsymbol{\Psi}_i(\hat{\alpha})\mathbf{y}_i \\
&= \mathbf{W}_i^{-1}\mathbf{X}_i^T\boldsymbol{\Psi}_i(\hat{\alpha})\mathbf{y}_i \\
&= \mathbf{W}_i^{-1}(\mathbf{1}_{n_i}^T\otimes\mathbf{B}^T)(\mathbf{I}-\mathbf{J}_{n_i}\otimes\mathbf{A}_{1i})\mathbf{y}_i \\
&= \mathbf{W}_i^{-1}\left\{(\mathbf{1}_{n_i}^T\otimes\mathbf{B}^T)-(n_i\mathbf{1}_{n_i}^T\otimes\mathbf{B}^T\mathbf{A}_{1i})\right\}\mathbf{y}_i \\
&= \mathbf{W}_i^{-1}\left\{\mathbf{1}_{n_i}^T\otimes\mathbf{B}^T(\mathbf{I}-n_i\mathbf{A}_{1i})\right\}\mathbf{y}_i \\
&= \left(\mathbf{1}_{n_i}^T\otimes\left\{\mathbf{W}_i^{-1}\mathbf{B}^T(\mathbf{I}-n_i\mathbf{A}_{1i})\right\}\right)\mathbf{y}_i \\
&= \mathbf{W}_i^{-1}\mathbf{B}^T(\mathbf{I}-n_i\mathbf{A}_{1i})\sum_{j=1}^{n_i}\mathbf{y}_{ij} \qquad \text{(A.8)} \\
&= \left(\mathbf{B}^T\left(\mathbf{I}_r-n_i\mathbf{A}_{1i}\right)\mathbf{B}\right)^{-1}\mathbf{B}^T\left(\mathbf{I}_r-n_i\mathbf{A}_{1i}\right)\bar{\mathbf{y}}_{i\cdot} \\
\hat{\mathbf{u}}_i &= \mathbf{Z}_i^T\boldsymbol{\Psi}_i(\hat{\alpha})(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_i)/\hat{\alpha} \\
&= (\mathbf{1}_{n_i}^T\otimes\mathbf{C}^T)(\mathbf{I}-\mathbf{J}_{n_i}\otimes\mathbf{A}_{1i})(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_i)/\hat{\alpha} \\
&= \left\{\mathbf{1}_{n_i}^T\otimes(\mathbf{C}^T-n_i\mathbf{C}^T\mathbf{A}_{1i})\right\}(\mathbf{y}_i-(\mathbf{1}_{n_i}\otimes\mathbf{B})\hat{\boldsymbol{\beta}}_i)/\hat{\alpha}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\hat{\alpha}}\mathbf{C}^T\left(\mathbf{I}_r - n_i\mathbf{A}_{1i}\right)\Big(\sum_{j=1}^{n_i}\mathbf{y}_{ij} - n_i\mathbf{B}\hat{\boldsymbol{\beta}}_i\Big)\\
&= \frac{n_i}{\hat{\alpha}}\mathbf{C}^T\left(\mathbf{I}_r - n_i\mathbf{A}_{1i}\right)\left(\bar{\mathbf{y}}_{i.} - \mathbf{B}\hat{\boldsymbol{\beta}}_i\right)
\end{aligned}
$$

(3.12): Before deriving the computation formula for the restricted log-likelihood from the already derived form (3.6) let us note that

$$
\begin{aligned}
&\{\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i(\alpha)\}^T\boldsymbol{\Psi}_i(\alpha)\{\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i(\alpha)\}\\
=\ &\{\mathbf{y}_i - (\mathbf{1}_{n_i}\otimes\mathbf{B})\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I} - \mathbf{J}_{n_i}\otimes\mathbf{A}_{1i})\{\mathbf{y}_i - (\mathbf{1}_{n_i}\otimes\mathbf{B})\boldsymbol{\beta}_i(\alpha)\}\\
=\ &\{\mathbf{y}_i - \mathbf{1}_{n_i}\otimes(\mathbf{B}\boldsymbol{\beta}_i(\alpha))\}^T\{\mathbf{y}_i - \mathbf{1}_{n_i}\otimes(\mathbf{B}\boldsymbol{\beta}_i(\alpha))\}\\
&-\{\mathbf{y}_i - \mathbf{1}_{n_i}\otimes(\mathbf{B}\boldsymbol{\beta}_i(\alpha))\}^T(\mathbf{J}_{n_i}\otimes\mathbf{A}_{1i})\{\mathbf{y}_i - \mathbf{1}_{n_i}\otimes(\mathbf{B}\boldsymbol{\beta}_i(\alpha))\}\\
=\ &\sum_{j=1}^{n_i}\{\mathbf{y}_{ij} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T\{\mathbf{y}_{ij} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}\\
&-\sum_{j=1}^{n_i}\left(\sum_{j=1}^{n_i}\{\mathbf{y}_{ij} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T\mathbf{A}_{1i}\right)\{\mathbf{y}_{ij} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}\\
=\ &\sum_{j=1}^{n_i}\{\mathbf{y}_{ij} - \bar{\mathbf{y}}_{\mathbf{i}\cdot} + \bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T\{\mathbf{y}_{ij} - \bar{\mathbf{y}}_{\mathbf{i}\cdot} + \bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}\\
&-n_i\{\bar{\mathbf{y}}_{i.} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(n_i\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{i.} - \mathbf{B}\boldsymbol{\beta}_i(\alpha)\}\\
=\ &\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})\\
&\qquad +n_i(\bar{\mathbf{y}}_{i.} - \mathbf{B}\boldsymbol{\beta}_i(\alpha))^T(\mathbf{I} - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - \mathbf{B}\boldsymbol{\beta}_i(\alpha))\\
=\ &\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})\\
&\qquad +n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})
\end{aligned}
\tag{A.9}
$$

where the last equality follows directly from $(A.8)$ and the definition of $\mathbf{A}_{2i}$. But this is precisely the definition of $S_i(\mathbf{y},\alpha)$ so we obtain the desired form of the log-likelihood:

$$
\begin{aligned}
l_R(\alpha) &= -\frac{n-p}{2}\log\sum_{i=1}^{g}\Big(\{\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}_i(\alpha)\}^T\boldsymbol{\Psi}_i(\alpha)\{\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}_i(\alpha)\}\Big)\Big/(n-p) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{I}+\mathbf{Z}_i^T\mathbf{Z}_i\mathbf{A}_i(\alpha)^{-1}| \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{X}_i^T\boldsymbol{\Psi}_i(\alpha)\mathbf{X}_i| \\
&= -\frac{n-p}{2}\log\sum_{i=1}^{g}\frac{1}{n-p}S_i(\mathbf{y},\alpha) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{I}+\frac{n_i}{\alpha}\mathbf{C}^T\mathbf{C}| \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\log|n_i\mathbf{B}^T(\mathbf{I}-n_i\mathbf{A}_{1i})\mathbf{B}| \qquad \text{(A.10)}
\end{aligned}
$$

(3.13): Note that the second part of the restricted log-likelihood ($A$.10) can be written in the following manner:

$$
-\frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{I}+\frac{n_i}{\alpha}\mathbf{C}^T\mathbf{C}| = -\frac{1}{2}\sum_{i=1}^{g}\left(\log|n_i\{\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I}\}| - K\log\alpha\right)
$$

Now using the chain rule ($A$.3) on the first part of the restricted log-likelihood and then using matrix rule ($A$.2) on the second and third part gives us the differentiated log-likelihood:

$$
\begin{aligned}
dl_R(\alpha) &= -\frac{n-p}{2}\left(\frac{\sum_{i=1}^{g}S_i'(\mathbf{y},\alpha)}{\sum_{i=1}^{g}S_i(\mathbf{y},\alpha)}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\left(\frac{1}{n_i}\mathrm{tr}\Big\{(\mathbf{C}^T\mathbf{C}+\frac{\alpha}{n_i}\mathbf{I})^{-1}\Big\}-\frac{K}{\alpha}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{g}\left(\mathrm{tr}\Big\{(\mathbf{W}_i)^{-1}\times d\mathbf{W}_i\Big\}\right) \qquad \text{(A.11)}
\end{aligned}
$$

where $S_i'(\mathbf{y},\alpha)$ is simply obtained by differentiating ($A$.9):

$$
\begin{aligned}
S_i'(\mathbf{y},\alpha) &= n_i(-n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})(\mathbf{I}-n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.}-n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
&\quad +n_i(\bar{\mathbf{y}}_{i.}-n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(-n_i d\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.}-n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})
\end{aligned}
$$

$$
\begin{aligned}
&+n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(-n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
= \;& -2n_i(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& -n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \qquad \text{(A.12)}
\end{aligned}
$$

The latter equality holds because the outcomes are scalars.

(3.14): Finally it is easily verified that the twice differentiated log-likelihood takes on the form

$$
\begin{aligned}
d^2 l_R(\alpha) \;=\; & -\frac{n-p}{2}\left(\frac{\left(\sum_{i=1}^g S_i''(\mathbf{y},\alpha)\right)\left(\sum_{i=1}^g S_i(\mathbf{y},\alpha)\right) - \left(\sum_{i=1}^g S_i'(\mathbf{y},\alpha)\right)^2}{\left(\sum_{i=1}^g S_i(\mathbf{y},\alpha)\right)^2}\right) \\
& -\frac{1}{2}\sum_{i=1}^g\left(-\frac{1}{n_i^2}\mathrm{tr}\Big\{(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{C} + \frac{\alpha}{n_i}\mathbf{I})^{-1}\Big\} + \frac{K}{\alpha^2}\right) \\
& -\frac{1}{2}\sum_{i=1}^g \mathrm{tr}\Big\{-\mathbf{W}_i^{-1}(d\mathbf{W}_i)\mathbf{W}_i^{-1}(d\mathbf{W}_i) + \mathbf{W}_i^{-1}(d^2\mathbf{W}_i)\Big\} \qquad \text{(A.13)}
\end{aligned}
$$

where $S_i''(\mathbf{y},\alpha)$ is obtained by differentiating the expression in $(A.12)$:

$$
\begin{aligned}
S_i''(\mathbf{y},\alpha) \;=\; & -2n_i\Big\{(n_i d^2\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& +(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(-n_i d\mathbf{A}_{1i})(\mathbf{y}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& +(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(-n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})\Big\} \\
& -n_i\Big\{(-n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d\mathbf{A}_{1i})(\mathbf{y}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& +(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d^2\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& +(\mathbf{y}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d\mathbf{A}_{1i})(-n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})\Big\} \\
= \;& -2n_i(n_i d^2\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& +4n_i(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d\mathbf{A}_{1i})(\mathbf{y}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& +2n_i(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(\mathbf{I} - n_i\mathbf{A}_{1i})(n_i d\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}) \\
& -n_i(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})^T(n_i d^2\mathbf{A}_{1i})(\bar{\mathbf{y}}_{i.} - n_i\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.})
\end{aligned}
$$

# Appendix B

In section 3.4 we compared the full model, (3.16), to the mean model, (3.17). We stated some consistency results and left them without proof. In this appendix we will prove the $\sqrt{r}$-consistency of the mean model REML estimate, $\hat{\sigma}^2_\varepsilon(\alpha)$, in (3.21) at the true value of $\alpha$. Then we will prove that the sample variance in (3.26) is $\sqrt{n_i}$-consistent. Finally we will prove that the full model REML estimate of the error variance, evaluated at the true $\alpha$, is consistent both as $r$ goes to infinity with $n_i$ fixed and as $n_i$ goes to infinity with $r$ fixed. For simplicity we will assume throughout that the number of curves, $n_i = n_0$, is constant across groups $i = 1, \ldots, g$.

(3.21): The REML estimator, of the error variance, from the mean profile mixed model, (3.19), evaluated at the true $\alpha$

$$\hat{\sigma}^2_\varepsilon(\alpha) = n_0\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}^*_i(\alpha)\}^T\boldsymbol{\Psi}^*_i(\alpha)\{\bar{\mathbf{y}}_{\mathbf{i}\cdot} - \mathbf{B}\boldsymbol{\beta}^*_i(\alpha)\}/(r-3), \tag{B.1}$$

is $\sqrt{r}$-consistent, as $r \to \infty$.

In order to prove this fact, we need the following lemmas

**Lemma B.0.1** *For a fixed $\alpha$, the matrix $(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T$, where $\mathbf{A}_{2i}$ is defined in* (3.11)*, is an idempotent matrix of rank $r-3$.*

*Proof:* Recalling the definitions of $\mathbf{A}_{2i}$ and $\mathbf{A}_{1i}$, from section 3.3, and noting that $\mathbf{A}_{1i}$ is symmetric, we get the desired result

$$\begin{aligned}
& (\mathbf{I}_r - n_0\mathbf{A}_{2i})^T(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T \\
= \;& \left\{\mathbf{I}_r - n_0(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T\right\}\left\{\mathbf{I}_r - n_0(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T\right\} \\
= \;& \left\{\mathbf{I}_r - n_0(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T\right\} \\
= \;& (\mathbf{I}_r - n_0\mathbf{A}_{2i})^T
\end{aligned}$$

Notice the cancelation of the matrix $n_0\mathbf{B}^T(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}$ and its inverse $\mathbf{W}_i^{-1}$ when multiplying the second terms in the curly brackets together.

Since the matrix $(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T$ is idempotent, the rank of the matrix is the same as the trace. Hence we obtain

$$\begin{aligned}
\text{rank}(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T &= \text{trace}(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T \\
&= \text{trace}(\mathbf{I}_r) - \text{trace}(n_0(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T) \\
&= r - \text{trace}(n_0\mathbf{B}^T(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}) \\
&= r - \text{trace}(\mathbf{I}_3) \\
&= r - 3
\end{aligned}$$

**Lemma B.0.2** *At the true $\alpha$, the random vector, $(\mathbf{I}_r - n_0\mathbf{A}_{2i})\bar{\mathbf{y}}_{i.}$, is normally distributed with mean $\mathbf{0}$ and covariance matrix*

$$\mathbf{\Sigma} = \mathbf{V}_i(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T,$$

*where $\mathbf{V}_i = Cov(\bar{\mathbf{y}}_{i.})$, as defined in section* 3.4.

*Proof:* The random vector $(\mathbf{I}_r - n_0\mathbf{A}_{2i})\bar{\mathbf{y}}_{i.}$ is obviously normal with mean $\mathbf{0}$. Before deriving the covariance matrix, recall that $(\sigma_\varepsilon^2/n_0)\mathbf{V}_i^{-1} = \mathbf{\Psi}_i^*(\alpha) = (\mathbf{I}_r - n_0\mathbf{A}_{1i})$.

$$\begin{aligned}
\mathbf{\Sigma} &= \text{Cov}\left((\mathbf{I}_r - n_0\mathbf{A}_{2i})\bar{\mathbf{y}}_{i.}\right) \\
&= (\mathbf{I}_r - n_0\mathbf{A}_{2i})\mathbf{V}_i(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T \\
&= (\mathbf{V}_i - n_0\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T(\mathbf{I} - n_0\mathbf{A}_{1i})\mathbf{V}_i)(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T \\
&= (\mathbf{V}_i - \sigma_\varepsilon^2\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T)(\mathbf{I}_r - n_0(\mathbf{I} - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T) \\
&= (\mathbf{V}_i - \sigma_\varepsilon^2\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T) \\
&= \mathbf{V}_i\left(\mathbf{I}_r - n_0(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}\mathbf{W}_i^{-1}\mathbf{B}^T\right) \\
&= \mathbf{V}_i(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T
\end{aligned}$$

Notice the cancelation of the matrix $n_0\mathbf{B}^T(\mathbf{I}_r - n_0\mathbf{A}_{1i})\mathbf{B}$ with its inverse $\mathbf{W}_i^{-1}$, in the fifth equality.

*Proof of* $(B.1)$: In section 3.4 we established that $\boldsymbol{\beta}_i^*(\alpha) = \boldsymbol{\beta}_i(\alpha)$, and in section 3.3 we established that $\mathbf{B}\boldsymbol{\beta}_i(\alpha) = n_0\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}$, hence the expression in $(B.1)$ can be written in the following way

$$\begin{aligned}
\hat{\sigma}_\varepsilon^2(\alpha) &= n_0\{\bar{\mathbf{y}}_{\mathbf{i}.} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}^T\boldsymbol{\Psi}_i^*(\alpha)\{\bar{\mathbf{y}}_{\mathbf{i}.} - \mathbf{B}\boldsymbol{\beta}_i^*(\alpha)\}/(r-3) \\
&= n_0\{\bar{\mathbf{y}}_{\mathbf{i}.} - n_0\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}\}^T\left[(\sigma_\varepsilon^2/n_0)\mathbf{V}_i^{-1}\right]\{\bar{\mathbf{y}}_{\mathbf{i}.} - n_0\mathbf{A}_{2i}\bar{\mathbf{y}}_{i.}\}/(r-3) \\
&= \sigma_\varepsilon^2\left\{\bar{\mathbf{y}}_{i.}^T(\mathbf{I}_r - n_0\mathbf{A}_{2i})^T\mathbf{V}_i^{-1}(\mathbf{I}_r - n_0\mathbf{A}_{2i})\bar{\mathbf{y}}_{i.}\right\}/(r-3) \\
&\stackrel{d}{=} \frac{\sigma_\varepsilon^2}{r-3}\chi_{r-3}^2,
\end{aligned}$$

since, by lemma $B.0.2$, $\mathbf{V}_i^{-1}\boldsymbol{\Sigma} = (\mathbf{I}_r - n_0\mathbf{A}_{2i})^T$, which is an idempotent matrix of rank $r-3$, by lemma $B.0.1$. Now we can use Chebyshev's inequality to prove the consistency result. Let $\delta > 0$, then

$$\begin{aligned}
P(|\hat{\sigma}_\varepsilon^2(\alpha) - \sigma_\varepsilon^2| > \delta) &= P\left(\left|\frac{\chi_{r-3}^2}{r-3} - 1\right| > \delta/\sigma_\varepsilon^2\right) \\
&\leq \left(\frac{\delta}{\sigma_\varepsilon^2}\right)^2 E\left[\left(\frac{\chi_{r-3}^2}{r-3} - 1\right)^2\right] \\
&= 2\left(\frac{\delta}{\sigma_\varepsilon^2}\right)^2\left(\frac{1}{r-3}\right)^2(r-3) \\
&= \left(\frac{\delta}{\sigma_\varepsilon^2}\right)^2\frac{2}{(r-3)} \xrightarrow{P} 0 \qquad \text{(B.2)}
\end{aligned}$$

as $r \to \infty$. It follows that $\hat{\sigma}_\varepsilon^2(\alpha)$, from $(B.1)$, is $\sqrt{r}$-consistent, as $r \to \infty$.

(3.26): The sample variance,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-p}\sum_{i=1}^{g}\sum_{j=1}^{n_0}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}) \qquad \text{(B.3)}$$

is a $\sqrt{n_0}$-consistent estimator of $\sigma_\varepsilon^2$, as $n_0 \to \infty$.

*Proof:*

$$\begin{aligned}
\hat{\sigma}_{\varepsilon}^2 &= \frac{1}{n-p}\sum_{i=1}^{g}\sum_{j=1}^{n_0}(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{i.}) \\
&= \frac{1}{n-p}\sum_{i=1}^{g}\sum_{j=1}^{n_0}(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_{i.})^T(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_{i.}) \\
&= \frac{1}{n-p}\sum_{i=1}^{g}\sum_{j=1}^{n_0}\sum_{k=1}^{r}(\boldsymbol{\varepsilon}_{ijk}-\bar{\boldsymbol{\varepsilon}}_{i.k})^2 \\
&\stackrel{d}{=} \frac{1}{n-p}\sum_{i=1}^{g}\sum_{k=1}^{r}\sigma_{\varepsilon}^2\chi^2_{n_0-1,i,k} \\
&= \frac{\sigma_{\varepsilon}^2}{grn_0-3g}\chi^2_{gr(n_0-1)} \\
&= \frac{1-\frac{1}{n_0}}{1-\frac{3g}{grn_0}}\left(\frac{\sigma_{\varepsilon}^2}{gr(n_0-1)}\chi^2_{gr(n_0-1)}\right) \qquad \text{(B.4)}
\end{aligned}$$

and it follows that $\hat{\sigma}_{\varepsilon}^2$, from $(B.3)$, is $\sqrt{n_0}$-consistent, as $n_0 \to \infty$.

Notice that if we keep $n_0$ fixed but let $r$ go to infinity we have the following

$$\hat{\sigma}_{\varepsilon}^2 \xrightarrow{P} (1-\frac{1}{n_0})\sigma_{\varepsilon}^2 \qquad \text{(B.5)}$$

Now it is easy to verify that at the true value of $\alpha$ the REML estimate of the error variance from the full model is consistent both as $n_0 \to \infty$ with $r$ fixed, and as $r \to \infty$ with $n_0$ fixed. It follows from the proof of the $\sqrt{r}$-consistency in $(B.1)$, that

$$\begin{aligned}
&\sum_{i=1}^{g}\frac{n_0}{n-p}\{\bar{\mathbf{y}}_{\mathbf{i}.}-\mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I}_r-n_0\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{\mathbf{i}.}-\mathbf{B}\boldsymbol{\beta}_i(\alpha)\} \qquad \text{(B.6)} \\
\stackrel{d}{=}\ & \frac{1}{n-p}\sum_{i=1}^{g}\sigma_{\varepsilon}^2\chi^2_{r-3} \\
=\ & \left(\frac{gr-p}{grn_0-p}\right)\frac{\sigma_{\varepsilon}^2}{gr-p}\chi^2_{gr-p} \\
=\ & \left(\frac{gr-p}{gr-p/n_0}\right)\frac{\sigma_{\varepsilon}^2/n_0}{gr-p}\chi^2_{gr-p}
\end{aligned}$$

$$\xrightarrow{P} \begin{cases} \sigma_\varepsilon^2/n_0 \text{ as } r \to \infty \\ 0 \text{ as } n_0 \to \infty \end{cases}$$

By the results in $(B.3)$, $(B.5)$ and $(B.6)$ it follows that the full model REML estimate is consistent both when $r$ goes to $\infty$ and when $n_0$ goes to $\infty$,

$$\begin{aligned} \hat{\sigma}_\varepsilon^2(\alpha) &= \frac{1}{n-p}\sum_{i=1}^{g}\sum_{j=1}^{n_0}(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{i.})^T(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{i.}) \\ &\quad + \sum_{i=1}^{g}\frac{n_0}{n-p}\{\bar{\mathbf{y}}_{\mathbf{i}\cdot}-\mathbf{B}\boldsymbol{\beta}_i(\alpha)\}^T(\mathbf{I}_r - n_0\mathbf{A}_{1i})\{\bar{\mathbf{y}}_{\mathbf{i}\cdot}-\mathbf{B}\boldsymbol{\beta}_i(\alpha)\} \\ &\xrightarrow{P} \begin{cases} (1-\frac{1}{n_0})\sigma_\varepsilon^2 + \sigma_\varepsilon^2/n_0 = \sigma_\varepsilon^2 & \text{as } r \to \infty \\ \sigma_\varepsilon^2 + 0 = \sigma_\varepsilon^2 & \text{as } n_0 \to \infty \end{cases} \end{aligned}$$

## BIBLIOGRAPHY


Brumback, B. A. and Rice, J. A. (1998), "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves," *Journal of the American Statistical Association*, 93, 961–976.

Brumback, B. A., Ruppert, D., and Wand, M. P. (1999), "Variable Selection and function estimation in additive nonparametric regression using a data-based prior: Comment," *Journal of the American Statistical Association*, 94, 794–797.

Crainiceanu, C. and Ruppert, D. (2004), "Likelihood ratio tests in linear mixed models with one variance component," *J. R. Statist. Soc. B*, 66, 165–185.

Crainiceanu, C., Ruppert, D., Claeskens, G., and Wand, M. P. (2005), "Exact Likelihood ratio tests for penalized splines," *Biometrika*, 92.

Cuevas, A., Febrero, M., and Fraiman, R. (2004), "An anova test for functional data," *Computational Statistics and Data Analysis*, 47, 111–122.

Eilers, P. H. C. and Marx, B. D. (1996), "Flexible smoothing with B-splines and penalties," *Statistical Science*, 11, 89–121.

Hastie, T. J. (1996), "Pseudosplines," *Journal of the Royal Statistical Society, Series B.*, 58, 379–396.

Henderson, C. (1950), "Estimation of genetic parameters," *Annals of Mathematical Statistics*, 21, 309–10.

Kimeldorf, G. S. and Wahba, G. (1970), "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines," *The Annals of Mathematical Statistics*, 41, 495–502.

Lindstrom, M. J. and Bates, D. M. (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *Journal of the American Statistical Association*, 83.

Moser, B. K. (1996), *Linear Models: A Mean Model Approach*, Academic Press.

Patterson, H. D. and Thompson, R. (1971), "Recovery of Inter-Block Information when Block Sizes are Unequal," *Biometrika*, 58, 545–554.

Pinheiro, J. C. and Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, New York: Springer-Verlag.

Ramsay, J. O. and Silverman, B. W. (1997), *Functional Data Analysis*, Springer-Verlag.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis, Second Edition*, Springer-Verlag.

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric Regression*, Cambridge University Press.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

Silverman, B. W. (1985), "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting," *Journal of the Royal Statistical Society, Series B*, 47, 1–52.

Smyth, G. K. (1989), "Generalized Linear Models with Varying Dispersion," *Journal of the Royal Statistical Society. Series B (Methodological)*, 51, 47–60.

Speed, T. (1991), "That BLUP is a Good Thing: The Estimation of Random Effects, Comment," *Statistical Science*, 6.

Wahba, G. (1978), "Improper Priors, Spline Smoothing, and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society, Series B*, 40, 364–372.

Wahba, G. (1990), *Spline Models for Observational Data*, Society for industrial and applied mathematics.

Wang, Y. (1998), "Smoothing Spline Models with Correlated Random Errors," *Journal of the American Statistical Association*, 93, 341–348.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998), "Semiparametric Stochastic Mixed Models for Longitudinal Data," *Journal of the American Statistical Association*, 93, 710–719.

Zhang, D., Lin, X., and Sowers, M. (2000), "Semiparametric Regression for Periodic Longitudinal Hormone Data from Multiple Menstrual Cycles," *Biometrics*, 56, 31–39.